University of Siena – Department of Medical Biotechnologies

Doctorate in Genetics, Oncology and Clinical Medicine

(GenOMeC)

XXXV Cycle (2019-2022)

Coordinator: Prof. Francesca Ariani

Development of multimodal systems for monitoring paediatric brain disorders

PhD Candidate:
Eng. Lorenzo Frassineti

Supervisor:
Prof. Simone Furini

Tutor:
Prof. Claudia Manfredi

Reviewers:
Dr. Renaud Viellevoye
Prof. Fabrizio Esposito

Discipline code: ING-INF/06 Biomedical Engineering

Lorenzo
Frassineti
05.03.2023
17:33:49
GMT+01:00

DON'T PANIC

Anything can be real.
Every imaginable thing is happening somewhere
along the dimensional axis.

I love deadlines.
I love the whooshing noise they make as they go by.

Douglas Noël Adams (DNA)

Abstract

In the last years, artificial intelligence (AI) methods are extensively applied in several fields, including healthcare, with several applications to support diagnostic approaches or treatments. The research activities carried on during my PhD work have been devoted to the development of AI methods to support neonatologists and paediatric neurologists in the detection, characterization, and monitoring of brain disorders in paediatric subjects. Specifically, the PhD work was focused on the development of multimodal systems for: neonatal and absence seizure detection; quantitative characterization of the speech phenotype for some genetic syndromes; prediction of the neurodevelopmental scales in newborns with sepsis.

In the first part of this PhD work, absence seizure detectors have been developed both for online and offline applications based on Electroencephalographic (EEG) signals and sonification algorithms. Following the encouraging results obtained for absence seizures, first attempts were made to validate EEG-based Neonatal Seizure Detectors (NSDs), a still tricky and time-consuming issue in the clinical practice. Moreover, Heart rate variability (HRV) analysis was proposed as an alternative approach for the detection of neonatal seizures. Experimental results confirmed the involvement of the Autonomic Nervous System during or close to neonatal seizures. The comparison between EEG-based NSDs and HRV ones confirmed that the best approach to detect neonatal seizures is still the EEG. However, when EEG techniques are not available, the use of HRV-based NSDs could be a promising alternative.

In the second part of this PhD work, quantitative acoustical analysis has been applied to the definition of the speech phenotype for four genetic syndromes: Down, Noonan, Costello and Smith-Magenis. Preliminary results confirm that acoustical measures could add helpful information for several syndromes with well-known language/voice impairments. Being completely non-invasive, acoustical analysis and AI methods might significantly contribute to the clinical assessment of such pathologies, also after surgical, pharmacological or logopaedic treatments and for long-term monitoring of the acoustical characteristics of the voice of these subjects.

The last part of this PhD thesis exploits the possibility of forecasting neurodevelopmental scores in preterm newborns with and without sepsis. Using AI regression models, reliable results at different time steps of the follow-up were obtained, both with EEG and HRV features. The BAYLEY-III test was used to compute the scores in three different domains: cognitive, language and motor. Results suggest that both EEG and HRV quantitative analysis could be helpful for the clinical staff, identifying the newborns at risk of neurodevelopmental delays.

Summing up, this PhD thesis shows how AI methods could be a valid support to clinicians in neurological paediatrics. Several experimental results are presented, showing possible applications and factual integration between AI techniques and clinical knowledge and needs, providing novel solutions and tools to support the clinical staff in the detection and characterization of brain diseases in infants and children.

List of scientific publications

Peer-reviewed journal publications

1. Frassineti, L., Barba, C., Melani, F., Piras, F., Guerrini, R., Manfredi, C., 2019. Automatic detection and sonification of nonmotor generalized onset epileptic seizures: Preliminary results. Brain Research, Vol. 1721, p. 146341. https://doi.org/10.1016/j.brainres.2019.146341.
2. Frassineti, L., Parente, A., Manfredi, C., 2021. Multiparametric EEG analysis of brain network dynamics during neonatal seizures. Journal of Neuroscience Methods, Vol. 348, p. 109003. https://doi.org/10.1016/j.jneumeth.2020.109003
3. Frassineti, L., Lanatà, A., Olmi, B., Manfredi, C., 2021. Multiscale Entropy Analysis of Heart Rate Variability in Neonatal Patients with and without Seizures. Bioengineering, Vol. 8, Issue 9, p. 122. https://doi.org/10.3390/bioengineering8090122
4. Olmi, B., Frassineti, L., Lanata, A., Manfredi, C., 2021. Automatic Detection of Epileptic Seizures in Neonatal Intensive Care Units Through EEG, ECG and Video Recordings: A Survey. IEEE Access, Vol. 9, pp. 138174–138191. https://doi.org/10.1109/access.2021.3118227.
5. Olmi, B., Manfredi, C., Frassineti, L., Dani, C., Lori, S., Bertini, G., Cossu, C., Bastianelli, M., Gabbanini, S., Lanatà, A., 2022. Heart Rate Variability Analysis for Seizure Detection in Neonatal Intensive Care Units. Bioengineering, Vol. 9, Issue 4, p. 165. https://doi.org/10.3390/bioengineering9040165

Peer-reviewed conference proceedings

1. Frassineti, L., Guerrini, R., Barba, C., Melani, F., Piras, F., Manfredi, C., 2019. Sonification techniques applied to EEG signals of nonmotor generalized onset epileptic seizures. 11th International Workshop, Models and Analysis of Vocal Emissions for Biomedical Applications, December, 17-19, 2019, ISSN 2704-5846, pp. 257-260. doi: 10.36253/978-88-6453-961-4.
2. Frassineti, L., Ermini, D., Fabbri, R., Manfredi, C., 2020. Neonatal Seizures Detection using Stationary Wavelet Transform and Deep Neural Networks: Preliminary Results. In 20th IEEE Mediterranean Electrotechnical Conference, June 16-18, 2020, Palermo. doi: 10.1109/MELECON48756.2020.9140713.
3. Frassineti, L., Manfredi, C. , Olmi, B., Lanatà, A., 2021. A Generalized Linear Model for an ECG-based Neonatal Seizure Detector. In 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp. 471-474. doi: 10.1109/EMBC46164.2021.9630841.
4. Frassineti, L., Lanatà, A., Mandredi, C., 2021. HRV analysis: a non-invasive approach to discriminate between newborns with and without seizures. In 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2021, pp. 52-55. doi: 10.1109/EMBC46164.2021.9629741.
5. Frassineti, L., Zucconi, A., Calà, F., Sforza, E., Onesimo, R., Leoni, C., Rigante, M., Manfredi, C., Zampino, G., 2021 . Analysis of vocal patterns as a diagnostic tool in patients with genetic syndromes. In Models and analysis of vocal emissions for biomedical applications: 12th international workshop: December 14-16, 2021: Firenze, Italy (pp. 83-86). doi: http://digital.casalini.it/9788855184496
6. Frassineti, L., Manfredi, C., Ermini, D., Fabbri, R., Olmi, B., Lanatà, A., 2022. Analysis of Brain-Heart Interactions in newborns with and without seizures using the Convergent Cross Mapping

approach. In 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp. 36-39, doi: 10.1109/EMBC48229.2022.9871141.
7. Olmi, B., Manfredi, C., Frassineti, L., Dani, C., Lori, S., Bertini, G., Gabbanini, S., Lanatà, A., 2022. Aggregate Channel Features for newborn face detection in Neonatal Intensive Care Units. In 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp. 455-458, doi: 10.1109/EMBC48229.2022.9871399.
8. Frassineti, L., Olmi, B., Lanatà, A., Lori, S., Gabbanini, S., Dani, C., Bertini, G., Cossu, C., Bastianelli, M. and Manfredi C., (2022, accepted). Artificial Intelligence in Neonatal Intensive Care Units: a multimodal approach for seizure detection and aetiology characterization. In 66° Congresso Nazionale Società Italiana di Neurofisiologia Clinica, May 18-21, 2021, Palermo, Italia – Neurogical Sciences.

List of training and teaching activities

Reviewer for peer-reviewed journals:

- Biomedical Signal Processing and Control ELSEVIER (5 papers)
- Scientific Reports NATURE (1 paper)
- Expert Systems With Applications ELSEVIER (1 paper)
- Early Human Development ELSEVIER (1 paper)
- IEEE Access, IEEE (1 paper)
- BMC Pediatrics, Springer (1 paper)
- Intelligent Medicine, ELSEVIER (1 paper)

Training at the University of Siena:

- MYCAREER - Introduction to EURAXESS, MSCA, ERC and Project Design, curatorship: Prof. Lorenzo Zanni.

Participation and cooperation in International and Italian Congresses:

- 11th International Workshop MAVEBA 2019 - Models and Analysis of Vocal Emissions for Biomedical Applications, Università degli Studi di Firenze (Firenze, Italy, December 17-19, 2019). Technical and scientific collaboration. Organizer and Chair: Prof. Claudia Manfredi.
- MELECON2020 IEEE Mediterranean Electrotechnical Conference (Virtual Conference, Palermo June 16-18, 2020).
- XXXIX Scuola GNB, "AI-enabled health care: from decision support to autonomous robots" (Virtual School, Bressanone, September 7-10, 2020).
- Second International Summer School on Technologies and Signal Processing in Perinatal Medicine (Virtual Conference, Pula-Cagliari, Italy, July 16-23, 2021). Organizer and Chair: Prof. Danilo Pani.
- 43rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society – Mexico (virtual conference November 1-5, 2021).
- 12th International Workshop MAVEBA 2021 - Models and Analysis of Vocal Emissions for Biomedical Applications, Università degli Studi di Firenze (Firenze, Italy, December 14-16, 2021). Technical and scientific collaboration. Organizer and Chair: Prof. Claudia Manfredi.
- 44th Annual International Conference of the IEEE Engineering in Medicine and Biology Society – Glasgow, Scotland, UK (July 11-15, 2022).
- 66th national congress SINC (Palermo, Italy, May 18-21, 2022).

Training and teaching activities outside the University of Siena:

- Trainee from March 2022 to June 2022 at **SmarTex s.r.l.** (via Giuntini 13L, Polo Scientifico e Tecnologico, Navacchio, Pisa, Italy), tutor: Dr. Rita Paradiso. Mainly activities: development of signal processing and artificial intelligence methods for apnoea detection and characterization by ECG and EDR signals.

- Co-supervisor of n. 1 Bachelor's Degree Thesis, for the School of Engineering, Curricula Electronics and Telecommunications Engineering, Università degli Studi di Firenze, Firenze, Italy:
    - Candidate Dr. Valentina Guarguagli, Thesis title: *Music Therapy with neuro-rehabilitative application: development of an experimental protocol*. (date of discussion: 04/12/2022, Supervisors: Prof. Claudia Manfredi, Prof. Antonio Lanatà).
- Co-supervisor of n. 3 Master's Degree Thesis, for the School of Engineering, Curricula Biomedical Engineering, Università degli Studi di Firenze, Firenze, Italy:
    - Candidate Eng. Benedetta Olmi, Thesis title: *Systems for detection and analysis of epileptic seizures in neonatal intensive care Units*. (date of discussion: 04/27/2021, Supervisors: Prof. Claudia Manfredi, Prof. Antonio Lanatà)
    - Candidate Dr. Federico Calà, Thesis title: *AI techniques for the acoustic characterization of genetic syndromes*. (date of discussion: 04/11/2022, Supervisors: Prof. Claudia Manfredi, Prof. Antonio Lanatà)
    - Candidate Dr. Angela Parente, Thesis title: *Machine learning techniques for studying the influence of sepsis on neurodevelopment in preterm infants: application to the Neonatal Intensive Care Unit AOU Careggi, Florence.* (date of discussion: 07/21/2022, Supervisors: Prof. Claudia Manfredi, Prof. Antonio Lanatà)
- Collaboration to teaching for the following courses:
    - **Biomedical Signal Processing** (Prof. Claudia Manfredi), Department of Information Engineering, School of Engineering, Università degli Studi di Firenze, Firenze, Italy. From 2018 to 2022.
    - **Rehabilitation Bioengineering** (Prof. Claudia Manfredi and Prof. Antonio Lanatà), Department of Information Engineering, School of Engineering, Università degli Studi di Firenze, Firenze, Italy. From 2018 to 2022

Table of contents

Introduction

In the last years, artificial intelligence (AI) and machine-learning (ML) models are extensively applied in several fields, including healthcare, where AI methods have found several applications to support diagnostic approaches or treatments. Most applications of AI and ML techniques concern adults and children, as a support to the clinical staff and patients in their daily activities. In the recent challenging era, due to COVID-19 outbreak, the use of telemedicine and AI approaches has provided a noteworthy breakthrough for healthcare, for a timely monitoring and feedback between the clinical staff and patients.

In this context, the research activities of this PhD work have been devoted in the development of AI methods to support neonatologists and paediatric neurologists in the detection, characterization, and monitoring of brain disorders in paediatric subjects. Indeed, AI applications for paediatric subjects are less investigated in literature than for adults, therefore several AI and ML solutions have still to be discovered or demonstrated as helpful for paediatrics subjects.

In the first part of this PhD thesis, the possibility to develop seizure detectors for paediatric subjects has been discussed and presented. Firstly, absence seizure detectors have been developed both for online and offline applications based on Electroencephalographic (EEG) signals. Such methods were validated on a private dataset collected at the A. Meyer Hospital, (Firenze, Italy), made up of 24 children with absence seizures. Using EEG entropy, wavelet and coherence features, Support Vector Machine (SVM) models gave an average $F1_{score}$ of 78% for offline mode on long-term recordings. Moreover, combined with a sonification algorithm, in online mode the same models gave an average $F1_{score}$ of 69%. Results confirmed that sonification algorithms and AI models could be used to support the clinical staff in the automatic detection of ictal events for absence seizures, as well as for monitoring their time evolution.

Following the encouraging results obtained for absence seizures, AI and ML models have been considered to develop reliable Neonatal Seizure Detectors (NSDs). Neonatal seizures are one of the most common neurological emergencies in Neonatal Intensive Care Units (NICUs). If not treated promptly they could have a negative impact on the neurodevelopment of the newborn. However, their detection is still tricky and time-consuming. The state-of-the-art of NSD has been analysed in this part of the PhD work, showing that the neonatal seizure detection is still an open issue. To address this problem, a Deep-Learning EEG-based NSD has been developed applying a combination of Stationary Wavelet Transform (SWT) and Fully Convolutional Neural Network (FCN). Methods were validated on 79 EEG signals (39 with seizure events) from a public dataset collected at the NICU of the Helsinki University Hospital (Helsinki, Finland). With the Leave-One-Subject-Out (LOSO) validation the SWT+FCN model gave a $F1_{score}$ of 48% and an Area Under ROC Curve (AUC) of 80%, showing that the neonatal seizure detection task is more complex than the absence seizure detection one. Moreover, the review of the literature suggested that the physiological mechanisms behind neonatal seizures are still unclear for some aetiologies, especially as far as the Central Nervous System (CNS) and its interactions with the Autonomic Nervous System (ANS) are concerned.

For these reasons, during this PhD project a more specific analysis of the brain network dynamics of neonatal seizures was performed, based on the recent evidence in the literature, that seizures and epilepsy could be a brain network disease. Circular Omega Complexity (COC) index and the Synchronizability (S) index were applied to the EEG signals from the Helsinki Dataset. COC and S describe the phase synchrony of the EEG signals and the degree of stability of the EEG system's globally synchronized state, respectively. Pre- post- and ictal periods were characterized by the two

indexes and they were compared to seizure-free signals. COC showed significant differences between seizure and seizure-free events (Mann-Whitney test, p-value <0.001 and Cohen's d 0.86). The combination of S and COC in standardized temporal instants provided a reliable description of the physiological behaviour of the brain network during and close to neonatal seizures. Findings confirm the usefulness of the evaluation of brain network dynamics over time for a better understanding and interpretation of the complex mechanisms behind neonatal seizures. The proposed methods could also support existing seizure detectors as a post-processing step in doubtful cases.

Finally, other source of information than EEG has been investigated to detect and characterize neonatal seizures. In fact, in the literature, it has been argued that also the ANS may be directly/indirectly involved during or close to a neonatal seizure. Thus, variations on the ANS dynamics might be measured to detect or characterize ictal events in newborns. Firstly, it has been evaluated if multiscale heart rate variability (HRV) entropy features could be used to discriminate between newborns with seizures and seizure-free ones. A cohort of 52 patients (33 with seizures) from the Helsinki dataset has been considered. Multiscale sample and fuzzy entropy showed significant differences between the two groups (Mann-Whitney Test p-values 0.02 and 0.008, respectively). Moreover, interictal activity showed significant differences between seizure events and seizure-free ones (Kruskal-Wallis Test, Bonferroni multiple-comparison post hoc correction, p-value 0.04 for multiscale sample entropy). These results suggested that such features could be used as inputs to NSDs. To this aim, several HRV-based NSDs have been developed during the PhD studies, using both multiscale entropy features and classical time- and frequency-domain HRV features. Models were validated on two datasets: the Helsinki Dataset and the Careggi dataset, collected at the Careggi University Hospital, Firenze, Italy, made of 51 patients, 22 with seizures. Among the ML models developed, a Patient Discriminant (PD) approach was proposed. PD assigns an index of seizure risk, identifying newborns with high risk of seizures. Encouraging results are achieved using a linear SVM, obtaining about 87% AUC and 89% for $F1_{score}$. Furthermore, using HRV-based NSD, AUC of 69% and 62% were obtained on the Helsinki and the Careggi datasets, respectively. These results were achieved applying a generalized linear model (GLM) for the Helsinki dataset, and a Gaussian SVM for the Careggi dataset. These findings confirm that neonatal seizures can alter the ANS of the newborn, and these changes could be detected through such novel approaches based on analysis techniques that make use of HRV measures, thus providing an aid to clinicians when EEG is not available. Finally, a first evaluation regarding the Brain-Heart Interactions (BHIs) analysis on neonatal seizures is presented. The BHIs were quantified by the Convergent Cross Mapping (CCM) approach, considering EEG and HRV signals from the Helsinki dataset. Preliminary results show that newborns with seizures have a lower degree of interaction between the CNS and the ANS than seizure-free ones (Mann Whitney test: p-value 0.02).

In the second part of this PhD thesis, two more topics regarding AI applications for paediatric subjects are discussed. The former is the quantitative acoustical analysis as a support for the characterization of the speech phenotype for some genetic syndromes. The latter is the prediction of neurodevelopmental scores in newborns with sepsis using AI methods.

Regarding the quantitative acoustical analysis on genetic syndromes, the analysis of audio and voice signals of subjects with genetic syndromes has been addressed. This evaluation was performed towards the definition of speech phenotype for four genetic syndromes: Down, Noonan, Costello and Smith-Magenis. Several acoustical features and multiscale entropy indexes were estimated using the BioVoice software tool and MATLAB routines. Differences between pathological and control cases and possible inter-syndromes differences have been evaluated with a non-invasive procedure based on the analysis of voice recorded with smartphones or microphones, both in hospital setting and at

home. This research was carried on taking into account two main aims: defining quantitative methods for the characterization of speech phenotype in genetic syndromes and providing methodologies available for the monitoring over time of the voice characteristics of the subject. Statistical results suggest that overall differences might exist between controls and subjects with genetic syndromes. Preliminary results confirmed the usefulness of both acoustical analysis and entropy techniques for the analyzed syndromes: indeed for some syndromes, a specific speech phenotype exists that might support the clinician, highlighting syndrome's characteristics not yet exploited.

AI regression models were evaluated for the analysis on newborns with sepsis, to predict the neurodevelopmental scores of preterm newborns with and without sepsis. The BAYLEY-III test was used to compute the scores in three different domains: cognitive, language and motor. The quantitative analysis was performed on EEG and ECG recordings acquired when the preterm infants reached the term age. The ECG cohort was made up of 48 preterm newborns, 27 of which with sepsis, while the EEG cohort was made up of 64 preterm newborns (38 with sepsis). Results are encouraging, giving a Mean Absolute Error (MAE) lower than 5 points for the BAYLEY-III cognitive and language scales at 6- and 12-months for ECG cohort. For the EEG cohort, a MAE lower than 5 points was obtained for the language scores at 6-months. Results suggest that an EEG or ECG exam could be a support to clinical staff, identifying the newborns at risk of neurodevelopmental delays.

In conclusion, this PhD thesis shows how AI methods could be a valid support in neurological paediatric applications. Several experimental results are presented related to childhood and neonatal seizures, quantitative language characteristics of genetic syndromes, neurodevelopmental delays, and neonatal sepsis. These results show only some of the possible applications and factual integration between AI techniques and clinical knowledge and needs, providing novel solutions and tools to support the clinical staff in the detection and characterization of brain diseases in infants and children.

List of Acronyms

ACC	Accuracy
AE, ApEn	Approximate Entropy
AED	anti-epileptic drug
aEEG	amplitude EEG
AI	Artificial Intelligence
ANS	autonomic nervous system
ASI	Activation Synchrony index
aTBN	average time between each number
AUC	Area under the ROC Curve
AUC_{cc}	Concatenated AUC
BHI	brain-heart interactions
BSI	Brain Synchrony Index
CAN	Central Autonomic Network
CCM	convergent cross mapping
CD	Careggi Dataset
CdLS	Cornelia de Lange syndrome
CDSA	Color Spectral Density Array
CI	Complexity Index
CNN	Convolutional Neural Network
CNS	central nervous system
COC	Circular Omega Complexity
CoS	Costello Syndrome
CS	Control Subject
DE	Distribution Entropy
DL	Deep-Learning
DS	Down Syndrome
ECG	electrocardiogram, electrocardiography
ECP	electroclinical seizures
EEG	electroencephalography, electroencephalogram
EGP	electrographic-only seizures
EMG	electromyography

EOS	Early Onset Sepsis
FCN	Fully Convolutional Neural Network
FDH, FD/h	False Detection per Hour
FDR	False Discovery Rate
FE	Fuzzy Entropy
FPR	False Positive Rate
GA	gestational age
GDR	Good Detection Rate
GSE	Generalized Sample Entropy
HD	Helsinki Dataset
HF	High Frequency
HIE	Hypoxic-Ischemic Encephalopathy
HR	heart rate
HRV	Heart Rate Variability
HTS	Hyper-Torus Synchrony
IBI	inter-beat-interval
ILAE	International League Against Epilepsy
kNN, KNN	k-nearest neighbours
KW	Kruskal-Wallis test
LF	Low Frequency
LOS	Late Onset Sepsis
LOSO	Leave-One-Subject-Out
MAE	Mean Absolute Error
MCC	Matthews Correlation Coefficient
MDE	Multiscale Distribution Entropy
MFDD	Mean False Detection Duration
MFE	Multiscale Fuzzy Entropy
ML	Machine-Learning
mRMR	Minimal-Redundancy-Maximal-Relevance
MSE	Mean Squared Error
MSE	Multiscale Sample Entropy
MW	Mann-Whitney test
NICU	Neonatal Intensive Care Unit

NPV	Negative Predictive Value
NS	Noonan Syndrome
NSD	Neonatal Seizure Detector
PA	Paediatric Age
PE	Permutation Entropy
PPV	Positive Predictive Value
PUVS	percentage of unvoiced segments
RMSSD	root mean square of successive RR interval differences
ROC	Receiver Operator Characteristic
S	Synchronizability
SE, SampEn	Sample Entropy
SEN	Sensitivity
SMS	Smith-Magenis Syndrome
SPE	Specificity
sTBN	standard deviation time between each number
SVM	Support Vector Machine
SVR	Support Vector Regression
SWT	Stationary Wavelet Transform
TRI	Triangular Index
VLF	Very Low Frequency
VSL	voice segment length

1. Automatic seizure detection in paediatric subjects: from absence seizures to neonatal seizures.

Some contents in this chapter are based on the following publications:

- *Frassineti, L., Barba, C., Melani, F., Piras, F., Guerrini, R., Manfredi, C., 2019. Automatic detection and sonification of nonmotor generalized onset epileptic seizures: Preliminary results. Brain Research, Vol. 1721, p. 146341. https://doi.org/10.1016/j.brainres.2019.146341.*
- *Frassineti, L., Guerrini, R., Barba, C., Melani, F., Piras, F., & Manfredi, C., 2019. Sonification techniques applied to EEG signals of nonmotor generalized onset epileptic seizures. 11th International Workshop, Models and Analysis of Vocal Emissions for Biomedical Applications, December, 17-19, 2019, Firenze, Italy, ISSN 2704-5846, pp. 257-260. doi: 10.36253/978-88-6453-961-4.*

In this chapter an introduction to the automatic seizure detection in paediatric subjects is provided. The main concepts related to childhood epilepsy, seizure detectors and the best practice for the performance evaluation of the detectors will be presented. The main findings, along with validation of the results, are described concerning a cohort of 24 young subjects affected by the so called "absence seizures" [Frassineti et al., 2019]. Several concepts introduced in this chapter will be used in the following sections of this thesis, for a better understanding of the main challenges and issues on the seizure detection field. Specifically, the automatic detection of absence seizures, the methods and terms used will introduce to a more complex problem in the seizure field: the recognition and characterization of neonatal seizures that will be discussed from Chapter 2 to Chapter 5.

This chapter is organized as follows: in Section 1.1 a brief introduction of what absence seizures are and the main results obtained in literature regarding their automatic detection will be discussed. In Section 1.2 the cohort considered will be presented. Section 1.3 and 1.4 will be devoted to present the methods and the main results. Discussion and conclusions are drawn in Section 1.5, showing how these findings may be helpful in the field on neonatal seizure detection.

1.1. Absence seizures and their automatic detection

The 1% of worldwide population suffers from epilepsy and these subjects are mainly present in childhood and in elderly age. Moreover, this incidence place epilepsy as one most common disorder in paediatric subjects [Zeng et al., 2016], while for adults only cerebrovascular diseases have a higher incidence. In the last years, epilepsy and epileptic seizures were classified in several ways and currently International League Against Epilepsy (ILAE) is updating such definitions in order to cover all the possible aspects of such disorders. Indeed nowadays, the pathophysiological mechanisms behind several kind of epilepsy and seizures are well understood, although for some of them it is still not clear the aetiology behind them. Thus, ILAE in 2017 provided an updated classification of seizures and epilepsy as shown in Figure 1.1 [Fisher et al., 2017].

As shown in Figure 1.1, among possible seizures there are the non-motor seizures, also known as absence seizures (or "*petit-mal*"). They are a typical disorder of the childhood, usually in children from 4 to 18 years old. Absence seizures represent the 10-17% of all the epileptic cases for such age range [Zeng et al., 2016]. They are characterized by seizure events of short duration (few seconds),

low motor component and high frequency of ictal events per day. The ictal event consists in a sudden interruption of awareness, unresponsiveness and blank stare. Sometimes it is possible to observe fine movements of the eyes and of the hands. If frequent, absences seizures could alter the learning skills, language development and socio-psychological attitude of the child [Verrotti et al., 2015]. Therefore, an early diagnosis of the disease is of utmost importance. Usually, absence seizures are benign disorders and often they tend to decrease in quantity and frequency at the end of the childhood. Moreover, anti-epileptic drugs (AEDs) can manage and mitigate the seizure burden in the child. That is true for the typical absences, while for other variants such as atypical or myoclonic it is not so. Moreover, the occurrence of absence seizure does not exclude the presence of other kind of epileptic seizures. As an example, the Lennox-Gastaut syndrome can be associated to the presence of atypical seizures [Ropper et al., 2005]. Therefore, not only the detection of seizures has to be done timely to improve the prognosis, but also the discrimination and identification of possible variants is crucial in order to find the best treatment for the child.

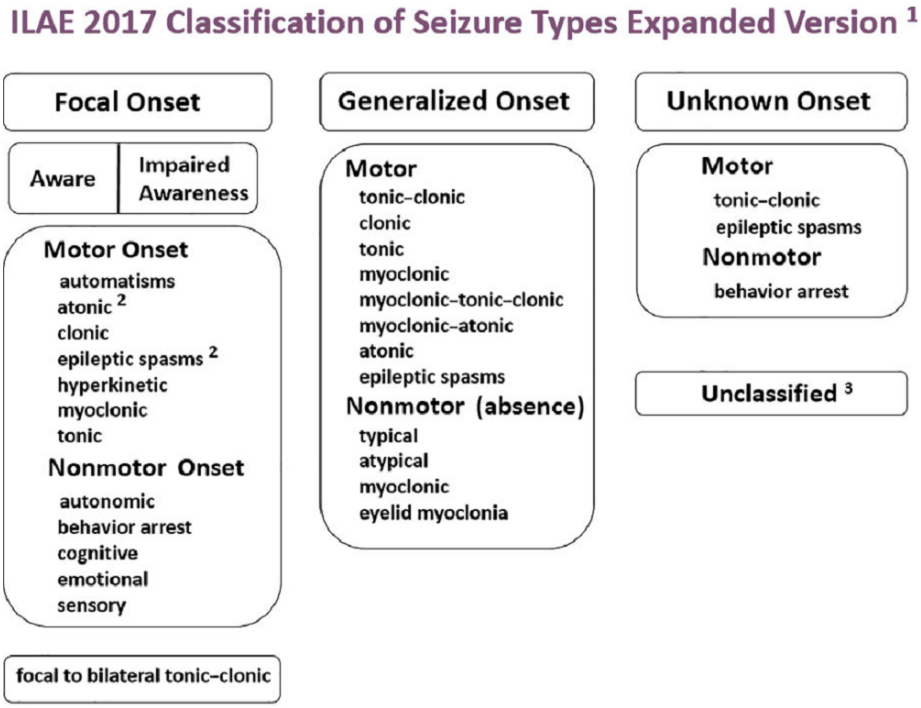


Figure 1.1. ILAE 2017 Classification of Seizure Types. [Fisher et al., 2017].

To diagnose epilepsy and the kind of seizures the gold-standard is electroencephalography (EEG) [Fisher et al., 2017]. Especially in children long-term video-EEG monitoring has improved diagnosis and treatment of epilepsy. As an example, in typical absences, EEG shows generalized spike-wave discharges at a frequency higher than 2.5 Hz. Conversely, atypical absences are characterized by a more gradual onset or termination or significant changes in tone supported by atypical, usually slow, generalized spike and wave discharges, at less than 2.5 Hz, in the EEG [Fisher et al., 2017].

However, for the neurologist, the analysis of long EEG recordings is time-consuming and it is usually performed in off-line mode. Thus, in the last years there has been a growing interest in methods concerning automatic early seizure detection, as well as technologies able to provide alert or detection of the occurrence of seizure events in on-line mode or quasi-real time (i.e. very low latency detection time after the seizure's onset). Concerning the automatic detection of absence seizures several works have been proposed, showing, however, a poor balance in terms of accuracy and latency time when applied in online mode [Alotaiby et al., 2014].

This short premise introduces the aims of the work presented in this chapter were Artificial Intelligence (AI) methods have been developed in order to run both in offline and online mode as support to the clinical staff for an ease detection and characterization of absence seizures.

1.2 The dataset of Absence Seizures

It is noteworthy that in the field of seizure detection only few public datasets are available online [Olmi et al., 2021]. This is a well-known limit in this field of research, because it may not be easy to assess the reproducibility nor the comparison of AI methods proposed in the literature. Specifically, for neonatal seizures only a single public dataset is available [Stevenson et al., 2019]. For this reason, in this work we had to build a private dataset to develop and validate the seizure detectors made of data from subjects with absence seizures.

Data were collected retrospectively and consisted of 30 EEG recordings from a set of 24 children (16 females, mean age 8.7 ± 3.8 years). Subjects were evaluated at the Paediatric Neurology Unit, Children's Hospital A. Meyer, Firenze, Italy. All the clinical information is summarized in Table 1.1 [Frassineti et al., 2019].

Table 1.1. Information on the 24 patients included in the study. [Frassineti et al., 2019].

Patient code	Epilepsy classification	Absence classification	Age at seizure onset/age at last seizure (ys)	Age at EEG (ys)	Ictal EEG	AED tried	Seizure free at last FU (yes/ no)	AED withdrawal at last FU (yes/no)	Number of analyzed recordings
1	CAE	Typical	4.9/6	5	3 HZ SW	ESM, VPA	Yes	No	1
2	GGE	Typical	5/still present	13	3 HZ SW	LTG, PB, VPA, ESM, ZNS, LEV, TPM, steroids	No	No	1
3	GGE	Atypical	11/still present	13	Polyspikes followed by SW	ESM, VPA, PB, CLZ, CNZ, STP, RFN, LEV, steroids	No	No	1
4	GGE	Typical	5/still present	18	3 HZ SW	ESM	No	No	1
5	GGE	Typical	4.8/7.5	6	3 HZ SW	VPA, ESM	Yes	No	1
6	CAE	Typical	6/7	6	3 HZ SW	VPA, ESM	Yes	Yes	1
7	CAE	Typical	2/5	4	3 HZ SW	VPA, ESM, LTG	Yes	Yes	1
8	GGE	Typical	9.8/10.2	10	3 HZ SW	ESM	Yes	No	3
9	CAE	Typical	5/6	6	3 HZ SW	VPA	Yes	Yes	1
10	GGE	Atypical	6/still present	9	2.5 HZ SW	VPA, ESM	No	No	1
11	CAE	Typical	5.6/7.2	7	3 HZ SW	ESM, VPA	Yes	Yes	1
12	CAE	Typical	5.5/6	6	3 HZ SW	ESM	Yes	No	1
13	GGE	Typical	6.6/7.4	7	3 HZ SW	VPA	Yes	No	3
14	CAE	Typical	10/12	11	3 HZ SW	ESM	Yes	No	1
15	CAE	Typical	6.5/ 7.1	6.9	3 HZ SW	VPA	Yes	No	1
16	CAE	Typical	7.6/8.1	8	3 HZ SW	VPA	Yes	Yes	1
17	CAE	Typical	9/10	9	3 HZ SW	ESM	Yes	No	1
18	JAE	Typical	13/17	17	3 HZ SW	VPA	Yes	Yes	1
19	CAE	Typical	7.3/8	7.8	3 HZ SW	VPA, ESM	Yes	Yes	1
20	CAE	Typical	5/6	5	3 HZ SW	VPA, ESM	Yes	Yes	1
21	CAE	Typical	6.5/still present	7	3 HZ SW	VPA, ESM	No	No	1
22	GGE	Typical	5.5/7	7	3 HZ SW	VPA, ESM	Yes	No	3
23	CAE	Typical	4/4.2	4	3 HZ SW	ESM	Yes	No	1
24	CAE	Typical	4.5/5.1	5	3 HZ SW	ESM, VPA	Yes	No	1

CAE: childhood absence epilepsy; CLZ: clobazam; CNZ: clonazepam; ESM: ethosuximide; FU: follow-up; GGE: genetic generalized epilepsy; Hz: hertz; JAE: juvenile absence epilepsy; LTG: lamotrigine; PB: phenobarbital; RFN: rufinamide; STP: stiripentol; SW: spikes and waves; TPM: Topiramate; TT: treatment, VPA: valproate; YS: years; ZNS: zonisamide.

For a first assessment of the proposed AI methods, for each subject the EEG recordings were cut in segments with mean length 53 ± 18 seconds (hereinafter short-term recordings). EEG was recorded using 19 channels in a longitudinal bipolar montage, placed according to the international standard 10 –20 system, with a sample frequency of 256Hz, and pre-filtered in the band 1.6-32 Hz for visual analysis. Then the same AI methods were tested considering the whole duration of the EEG: 47 ± 10 minutes (hereinafter long-term recordings). EEG recordings were labelled by an expert neurologist

of the Hospital A. Meyer. For each channel the neurologist marked the onset and the offset of each seizure event.

1.3 Methods

All the proposed methods are implemented under the MATLAB software tool [MATLAB and Statistics and Machine Learning Toolbox, release 2017b] installed on a Hp Pavillion 15 notebook (OS Windows 10, 64 bit) Intel Core i7-5500U processor, CPU 2.40 GHz, RAM 16 Gb. In Figure 1.2 the workflow adopted for the development of the Absence Seizure Detectors is shown. For the Permutation Entropy (PE) the toolbox developed by Jesús Monge-Álvarez was used [Jesús Monge-Álvarez, 2023].

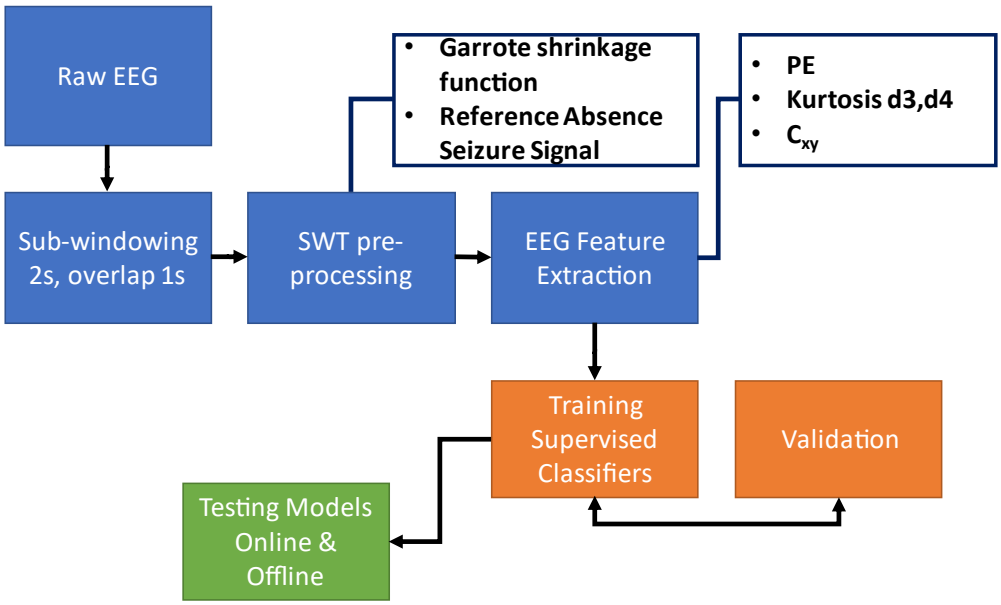


Figure 1.2. Workflow for the development of Absence Seizure Detectors.

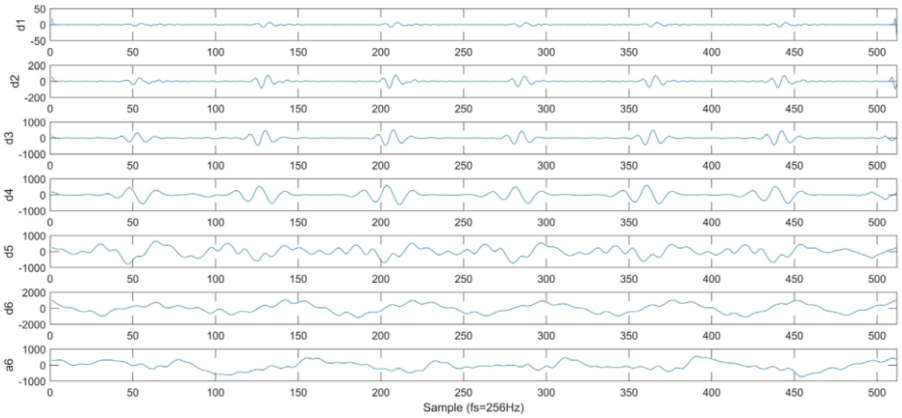
1.3.1 Pre-Processing and SWT filtering

It is well known that one of the biggest issue on scalp EEG is the presence of artifacts that may mask the presence of electrophysiological features of interest. Several methods were proposed in literature to mitigate or detect artifacts [Islam et al., 2016a]. In this study one of the aims was the development of both on-line and off-line analysis methods. Therefore, the following requirements [Patel et al., 2014], for artifact pre-processing had to be taken into account:

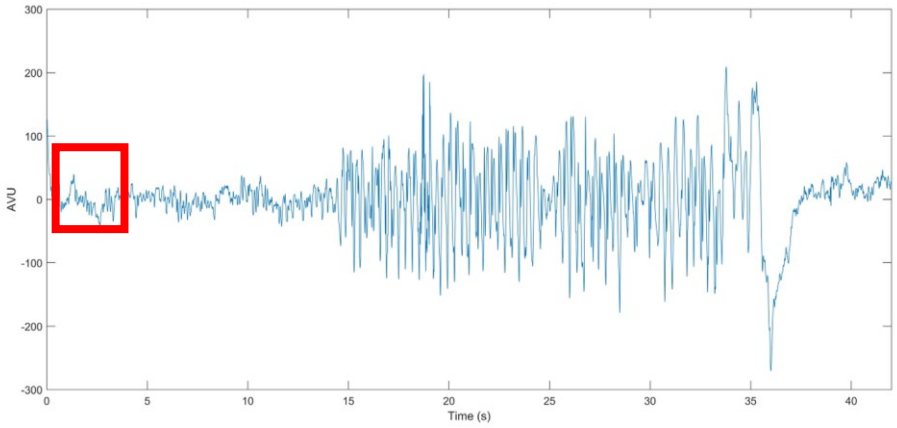
- Low computational time
- Only EEG signals available, no other physiological signals available such as electrocardiogram (ECG), or Electromyogram (EMG).
- Methods optimized for a specific clinical case: absence seizures.

Considering such premises for the pre-processing, we adapted the original method proposed by Md Kafiul Islam et al. [Islam et al., 2016b]. It applies the Stationary Wavelet Transform (SWT) [Nason and Silverman, 1997] to the original EEG time series. In our study EEG signals were divided into windows with length of 2 seconds, with overlap of 1 second. This operation is also called sub-

windowing procedure [Olmi et al., 2021]. An illustrative example of SWT decomposition on an EEG channel (C4-P4) is shown in Figure 1.3.



(a)



(b)

Figure 1.3. (a) The SWT decomposition of the first 2s epoch, the red square in (b), of the original raw EEG signal, channel C4-P4. [Frassineti et al., 2019].

After the selection of a fixed number of decomposition levels – 6 in our case – the method applied the following garrote shrinkage function $g(i,j)$ [Gao, 1998] as shown in Equation 1.1.

$$g(i,j) = \begin{cases} d_{i,j} & |d_{i,j}| \leq t_{i,j} \\ \frac{t_{i,j}}{d_{i,j}} & |d_{i,j}| > t_{i,j} \end{cases} \tag{1.1}$$

Where $d_{i,j}$ are the coefficients of the detail level, where i is the i-th EEG channel and j is the corresponding level of detail, that are filtered if larger than a given threshold $t_{i,j}$ that is also known as the modified universal threshold [Islam et al., 2016b], and is defined as follows:

$$t_{i,j} = K\alpha_{i,j}\sqrt{2\ln N} \tag{1.2}$$

Where $\alpha_{i,j}$ is defined as:

$$\alpha_{i,j} = \frac{median(|d_{i,j}|)}{0.6745} \tag{1.3}$$

The value of K in (1.2) varies according to the correlation index C_s. C_s is the absolute Pearson's correlation coefficient between the current signal analysed and a reference EEG signal of absence

seizure defined a priori for each given SWT decomposition level. An example of reference signal used in this work is shown in Figure 1.4.

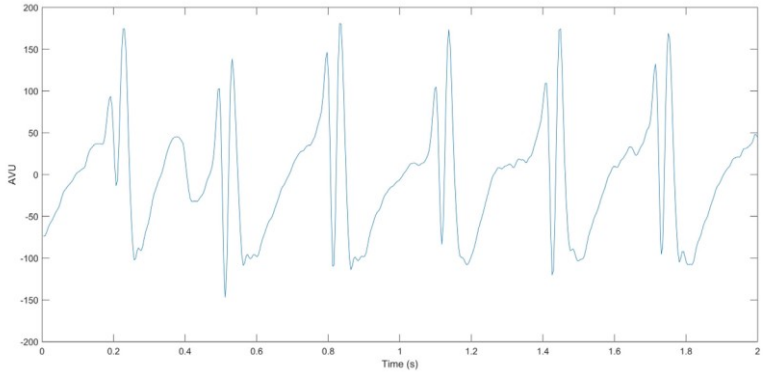


Figure 1.4. Example of EEG reference signal. [Frassineti et al., 2019].

Thus, as in [Islam et al., 2016b], K varies according to the following decision process:

- If $C_s \geq T_{high} \rightarrow no\ denoising$
- If $T_{low} \leq C_s < T_{high} \rightarrow K = 1.5$
- Otherwise K=1

The thresholds $T_{high} = 0.7$ and $T_{low} = 0.5$ were chosen according to [Islam et al., 2016b]. Instead, the number of 6 levels for the SWT decomposition was chosen because the detail coefficients d6 represent well the frequency range between 2Hz and 4Hz where the frequencies typical for absence are included [Frassineti et al., 2019]. Instead, the detail coefficients d1 and d2 as well as the approximation coefficients a6 were set to 0 as they do not carry useful information in the presented application. After the SWT filtering the EEG time series is reconstructed by the inverse operation ISWT. The filtering procedure was applied to all EEG channels. Regarding the mother-wavelet function used for SWT the following were evaluated: Biorthogonal 3.3 (Bior 3.3) [Upadhyay et al., 2016] and Haar [Islam et al., 2016b]. They were compared to the case when no filtering was applied. The results of this comparison are shown in section 1.4. Examples of the filtering procedure are reported in Figure 1.5.

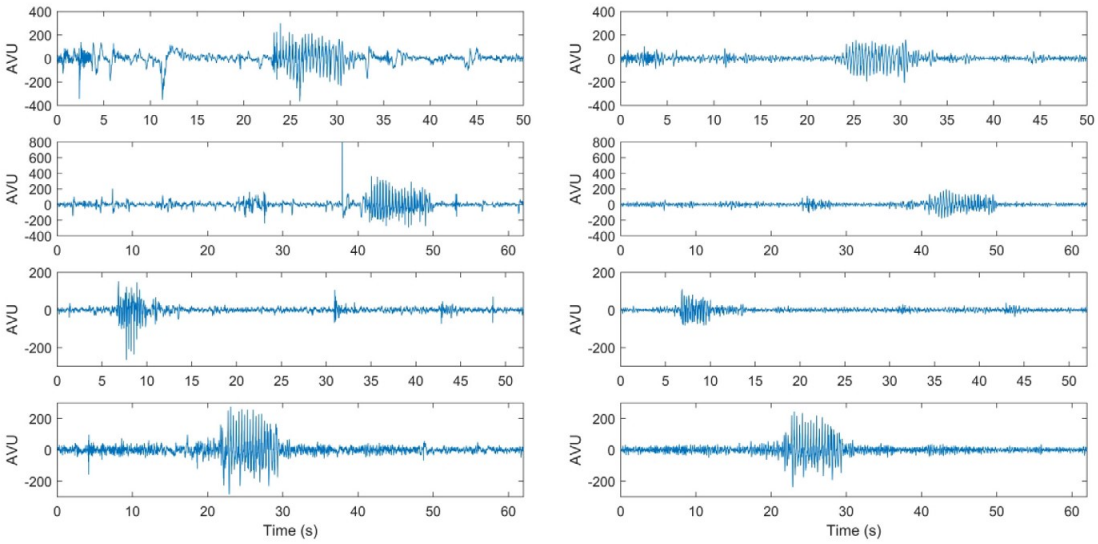


Figure 1.5. Example of the proposed SWT-based filtering. Left original EEG signals; right: filtered signals. [Frassineti et al., 2019].

1.3.2 EEG feature extraction

Providing an exhaustive list of quantitative EEG features developed in literature is out of the aim of this PhD thesis and for further details please see the dedicated reviews [Olmi et al., 2021, Baumgartner et al., 2018, Paul 2018, Acharya et al., 2013]. We point out that several EEG features proposed in literature were specific for the context analysed, and this is true also for the case of absence seizures. In this work both single channel metrics and multichannel metrics were evaluated [Frassineti et al., 2019].

Among the possible single-channel features available a specific domain has grown in interest in the field of seizure detection: the entropy-domain features. Entropy features for physiological time series provide a sort of measure of complexity and unpredictability, and in the last years it was proved that they are well suited for the detection and characterization of epileptic seizures [Upadhyay et al. 2016]. Among the entropy indexes one of them was evaluated in this work: the Permutation Entropy (PE) [Bandt and Pompe, 2002], as it has already shown promising results when applied to time series concerning absence seizures [Li et al., 2014].

Following the original definition of PE given by Bandt et al. [Bandt and Pompe, 2002], let N be the length in samples of the times series, from which the (N-m-1) delay vectors $X_t = [x_t, \dots, x_{t+(m-1)}]$ are built, where m is the embedding dimension. Then each vector X_t is rearranged in ascending order, let j the positional shift of each original element. Thus, for each delay vector there will be $m!$ re-ordered patterns also called motifs π_i, i=1…$m!$. The possible occurrences of the π i-th motif in the time series are indicated with $f(\pi_i)$, and the relative frequency is given by:

$$p(\pi_i) = \frac{f(\pi_i)}{N - m + 1} \tag{1.4}$$

Thus, it is possible to define the PE through the following equation:

$$PE := -\sum_{\pi_i}^{m!} p(\pi_i) \log_2 p(\pi_i) \tag{1.5}$$

According to [Bandt and Pompe, 2002] in this work m=3 and N=512. PE was computed for each EEG channel considered and for each window of 2 seconds.

Besides PE as single channel features the following measures were included in the analysis: the kurtosis of the SWT detail coefficients d3 and d4 [Upadhyay et al., 2016].

Finally, as multi-channel feature, the index of coherence across all EEG channels in the frequency range typical for absence seizures (i.e. between 2Hz and 4Hz) was added. The index of coherence provides a measure of phase synchronization between EEG signals [Aarabi et al., 2008]. In this work the Mean Magnitude squared coherence $C_{xy}(f)$ [Frassineti et al., 2019], described in Equation 1.6, was used:

$$C_{xy}(f) = \frac{|P_{xy}|^2}{P_{xx}(f)P_{yy}(f)} \tag{1.6}$$

Where $P_{xy}(f)$ is the cross power spectral density between each pair of derivations (x and y), calculated for frequencies between 2Hz and 4Hz, while $P_{xx}(f)$ and $P_{yy}(f)$ are the power spectral densities of the signals x and y in the same frequency range.

All the EEG derivations for C_{xy} were considered because absence seizures are considered as generalized seizure [Fisher et al., 2017], thus it is often possible to evaluate changes in synchronization across all the scalp of the subject. It is important to remark that for focal seizures, C_{xy} could not work without a previous selection of EEG channels that contain the seizure event.

In summary the following features from the EEG filtered time series were considered:

- PE (m=3) for all the 19 EEG derivations
- Kurtosis of detail coefficients d3 and d4 for all the 19 EEG derivations
- Mean Magnitude-squared coherence between 2Hz and 4Hz considering all the 19 EEG derivations.

Finally, all the features were normalized across subjects (between 0 and 1) and the outliers were detected and removed from the observations. In this work an observation was defined as outlier if one of its features is three times larger than the absolute deviation of the median MAD [Sachs and Berkovits, 1984].

1.3.3 Performance assessment of a seizure detector and choice of the supervised classifiers

Before starting to talk about the procedure adopted for the selection of the supervised classifiers as absence seizure detectors, it is important to introduce which metrics are often used to validate and assess the performance of a seizure detector.

Usually, a binary supervised classifier is evaluated by the Accuracy metric (ACC) (Equation 1.7), that provides a good synthesis regarding the classifier's performance for both the two classes considered. ACC can be defined by Equation (1.7):

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{1.7}$$

Often in a seizure detection problem the True Positive (TP) are the epochs (i.e. the EEG sub-windows) correctly classified as seizure epochs and the True Negative (TN) are the epochs correctly classified as windows without any seizure events [Olmi et al., 2021]. False Positive (FP) and False Negative (FN) are the epochs wrongly classified by the classifier with respect to the label provided by the ground truth, in our case the neurologist's labels.

Often the well-known confusion matrix is used to have a complete overview of the performance of a classifier in terms of sensitivity and specificity [Olmi et al., 2021]. Sensitivity (SEN) is defined as follows: $SEN = TP/(TP + FN)$, while Specificity (SPE) is: $TN/(TN + FP)$. However, the seizure detection problem for long-term EEG recordings can be considered as an unbalanced classification problem. In other words, the epochs with seizure events are much less than the interictal periods, therefore the ACC metric could mask the real performance of the detectors as far as the detection of seizure epochs are concerned. For these reasons it is a good practice to use other epoch-based metrics for the seizure detection problem to take into account the unbalance between the two classes. In this work the following metrics [Frassineti et al., 2019] were added:

$$Balanced\ Accuracy, BACC := \frac{\frac{TP}{TP + FN} + \frac{TN}{TN + FP}}{2} \tag{1.8}$$

$$F1_{score} := 2\frac{Precision\ \times Recall}{Precision\ + Recall} \tag{1.9}$$

$$Matthews\ Correlation\ Coefficient, \qquad MCC := \frac{TP \times TN - FP \times FN}{\sqrt{(TP+TN)(TP+FN)(TN+FP)(TN+FN)}} \tag{1.10}$$

For $F1_{score}$: Precision is defined as the percentage of correctly labelled seizure epochs, and Recall is the same as Sensitivity. Moreover, to evaluate the performance of the classifiers during online applications two more metrics were considered: one is the latency time given by the sum of the time to get new data (1s=overlap) plus the processing time of the method, according to Equation 1.11:

$$t_{latency} = t_{acquisition} + t_{processing} \tag{1.11}$$

The other one is the False Positive Rate (FPR), that evaluates the occurrence of misclassifications in the interictal phase for long-term recordings. FPR can be described as in Equation 1.12:

$$FPR = \frac{FP}{FP + TN} \tag{1.12}$$

During the validation of the classifiers, the best between them was identified as the one with the highest $F1_{score}$. As supervised models, both linear and non-linear Support Vector Machine (SVM), k-NN, Boosted Tree and Logistic Regression models were investigated. All these models were considered as they were already found useful in the seizure detection problem [Olmi et al., 2021, Baumgartner et al., 2018, Upadhyay et al., 2016]. Moreover, to avoid both overestimation of the classifier's performance and overfitting issues, all the classifiers were validated using a cross-validation procedure [Hastie et al., 2001]. For internal cross-validation 6 subjects were considered in order to find the best hyperparameters. This subset was selected a priori by the clinical staff. Then the remaining 24 EEG recordings were used as test set, computing all the metrics previously introduced in this subsection. Subjects with atypical absences were included only in the test set. The average processing time on the 24 test cases was 0.25s for each iteration. Therefore, the average latency time was almost 1.25s, compatible with online application. Finally, we remark that all the metrics described in this subsection are epoch-based. Moreover, we remark that also event-based and patient-based metrics were provided in literature to have a complete overview of the seizure detector's performance [Olmi et al., 2021]. In this work only epoch-based metrics were considered because they represent the first assessment for novel seizure detectors. We point out that EEG epochs from 2s to 4s at least are required for the neurologist to detect an absence seizure [Keilson et al., 1987]. Event- and patient-based are commonly used during the clinical validation of such methods, thus the results shown in section 1.4 must be considered as preliminary.

1.3.4 Post-processing and sonification

The last part of the implemented procedure was the application of spatial and temporal thresholds to the detection obtained by the supervised classifiers [Frassineti et al., 2019]. In this work, an epoch was classified as a seizure epoch if at least half plus one of the EEG channels were classified as seizure epoch at the same time. As for the temporal threshold a part of the signal was defined an absence (or a part of it) if and only if it is classified as absence for at least two seconds and separated each other by at least one second [Keilson et al., 1987]. Regarding the methods in offline mode, spatial and temporal thresholds were applied before the computation of the performance metrics. Moreover, at the end of the EEG examination a summary report was produced for each patient where the following information was reported:

- Number of absences found their duration and time occurrence.

- Average frequency, to discriminate between typical and atypical absence seizures.
- Derivation involved and duration of the involvement.

Instead for the online methods only the spatial threshold was considered. Moreover, to provide a sort of alert for the clinical staff of the occurrence of a seizure event in quasi-real time a sonification procedure was applied. The commonly accepted definition of sonification comes from G. Kramer et al. [Kramer et al., 1999]:

"*Sonification is the transformation of data relations into perceived relations in an acoustic signal for the purposes of facilitating communication or interpretation*".

There are several reasons why it may be useful to turn graphical and/or numerical information into sounds: overload of visual information; providing a support to visual information; speeding up the interpretation of information processes for real-time or online applications [Loui et al., 2014, Temko et al., 2015a]. For a correct sonification procedure, it is essential to develop a method that allows discriminating between the relevant information and artefacts in almost real time. In this work the following procedure was applied:

$$[\text{ABS-ABS-ABS}] \rightarrow [beep_1 beep_2 \rightarrow sound_{abs}]$$

In other words, if the automatic recognition detects three consecutive seconds as seizure epochs, the first two will produce 2 beeps, which allow a first discrimination between possible event of absence seizure and a false positive, while the third second and any following ones will produce a specific sound described in Equation 1.13.

$$sound_{abs} = \sum_{i=1}^{N} E_{sound_i} \sin(2\pi f_{osc} t) \tag{1.13}$$

Where E_{sound_i} are the oversampled coefficients of d6 (up to 48kHz, the playback audio frequency), and f_{osc} is the sound oscillation frequency set to the note C at 256Hz. The choice of 2 seconds played as beeps was defined since absences less than two seconds cannot be considered as possible seizures but just a sort of interictal activities.

1.4 Absence Seizure Detection: Results

In this subsection the main results obtained during the development and validation of the absence seizure detectors are reported. In Table 1.2 the results related to the comparison between the two different mother-wavelet functions and no-filtering on the short-term recordings are shown. The mean time duration was 53s ±18s. The supervised model was a Fine Gaussian SVM (FGSVM) [Burges, 1998].

Table 1.2. Results of the comparison on the short-time EEG recordings between different mother-wavelet functions (Bior 3.3 and Haar) and when no SWT filter was applied (No-Filter). Both mean and standard deviations are reported (μ±σ). [Frassineti et al., 2019].

	BACC %	**F1score %**	**MCC**
Bior 3.3	93±5	89±6	0.86±0.07
Haar	88±5	82±9	0.78±0.08
No-Filter	65±11	43±16	0.40±0.11

In Table 1.3 results obtained for all the supervised models both on the short-term time series and during validation are reported (without any post-processing). Instead on Table 1.4 the results obtained on the long-term time series are shown, using the FGSVM classifiers and applying the online and offline post-processing. The average time duration for the long-term recordings was 47±10 minutes.

Table 1.3. Results for the validation set on the short-time series, using different families of supervised classifiers. Both mean and the standard deviations are reported (μ±σ). [Frassineti et al., 2019].

	$F1_{score}$ Validation %	$F1_{score}$ %	MCC	BACC %
FGSVM	93	89±6	0.86±0.07	93±5
k-NN (k=1)	89	85±7	0.81±0.08	90±5
Boosted tree	92	89±6	0.86±0.06	92±5
Logistic Regression	90	87±6	0.83±0.08	91±5

Table 1.4. Results obtained on the test set of 24 long-term recordings. Both the performance without any post-processing and with online and offline methods are shown. Both mean and standard deviations are reported (μ±σ). [Frassineti et al., 2019].

	BACC %	$F1_{score}$ %	MCC	FPR %
FGSVM (Bior 3.3)	87±6	53±16	0.55±0.14	2.4±2.0
ONLINE	89±6	69±15	0.70±0.14	1.1±1.0
OFFLINE	91±5	78±15	0.78±0.14	0.8±1.0

1.5 Discussion and Conclusions

In this chapter a complete workflow for the development of an absence seizure detector both for online and offline mode was presented. From Table 1.2 it is possible to confirm that a pre-processing step is often required or at least recommended before starting the analysis on EEG signals at least in the case of absence seizure detection. While in this work only EEG signals were considered, other works have already proved that adding other physiological signals such as ECG or EMG can improve the detection of artefacts [Olmi et al., 2021]. Results confirmed, as in [Li et al., 2014], that entropy features are able to characterize absence seizures, and these features could be successfully used as input of supervised classifiers. Moreover, the use of multi-channel features such as C_{xy} for generalized seizures could be helpful in their characterization and detection [Mormann et al., 2000].

However, the features considered were extremely specific for absence seizures, therefore a fundamental step for each seizure detector that employs machine-learning methods will be the investigation and selection of specific features. This issue might be partially overcome by the use of deep-learning methods [Olmi et al., 2021]. In this work there is no mention to the possible methods available for the feature selection (e.g. mRMR [Ding and Peng, 2005], ReliefF [Robnik-Šikonja and Kononenko 2003]) or transformation (e.g. PCA, ICA [Hastie et al., 2001]). These aspects will be discussed in the next chapters of this PhD thesis. In this work the selection of the best set of features was made by a-priori clinical info and literature review. Regarding the performance assessment it is important to remark that epoch-based metrics should take into account the intrinsic imbalance between seizure events and interictal periods as already suggested by [Olmi et al. 2021, Temko et al., 2015a].

The results obtained in Table 1.4 for the offline mode, confirmed that for several kind of seizures in adults and children the problem of seizure detection may be almost solved when robust EEG analysis techniques are available [Chisci et al., 2010]. In fact, for the adults, already some tools include an off-line seizure detector among their utilities [Vidyaratne and Iftekharuddin, 2017]. The results showed in this chapter, though preliminary, suggest that also for children this solution could be adopted and integrated in clinical practice to support the clinical staff.

As shown in Table 1.3 and 1.4, the BACC parameter does not vary much between short-term and long-term analysis, while the $F1_{score}$ decreases. The PRECISION parameter in the F1 score formula (Equation 1.9) drastically drops due to the heavy imbalance in the number of test samples between seizures and interictal activity, which inevitably raises the number of false positives (also in case of high values of specificity, 98% average in this study). On the contrary, with the BACC parameter the unbalance has less influence on the test result because false positives are related only to true negatives. Possible improvements could be obtained adding more reference signals and finding other features for discrimination. For the online methods a sonification approach was implemented in order to alert the clinical staff recognizing the audio pattern associated to the absence seizures (two consecutive beeps and then the d6 sonified). This approach could e.g. be applied to an earphone, and might allow the clinical staff to concentrate their attention on the clinical signs during the ictal events, evaluating, as an example, the degree of awareness of the subject and, at the same time, monitoring the progression of the ictal events towards their offset (the end of sounds). To better explain the usefulness of EEG sonification, an example of a sonified EEG is shown in Figure 1.6 (channel C4-P4). The dashed rectangle represents the onset and offset of the seizure event labelled by the neurologist. The upper figure is the original EEG signals while the lower one shows the results of FGSVM detection and after the online post-processing, that is the sonified EEG (where Arbitrary Units AU are >0). From Figure 1.6 qualitatively shows that the sonified part significantly reduced the amount of signal to be analysed. Moreover, the adopted sonification scheme allows a discrimination between the seizure event and false positives (e.g. the continuous rectangles).

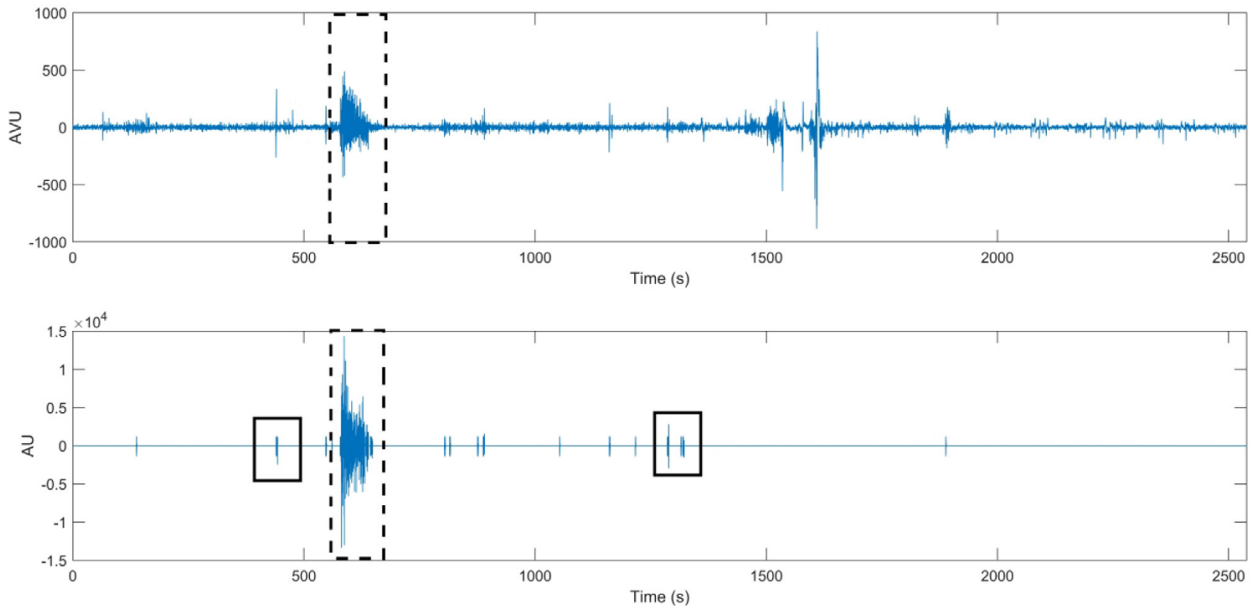


Figure 1.6. Example of EEG sonification. Above the EEG channel; below the sonified signal after the detection by FGSVM and online postprocessing. The dashed lines represent the onset and offset of the seizure event labelled by the neurologist. AVU= Arbitrary Voltage Units, AU= Arbitrary Units. [Frassineti et al., 2019]

The performance obtained applying the online and offline post-processing confirmed that the correct detection of the seizures is a crucial step in order to improve the accuracy of the AI systems, reducing the number of false positives (Table 1.4).

This work had some limits: no control cases were considered. Indeed, for a clinical validation of the methods a set of them should be defined. Although we obtained promising results even on subjects with atypical absences, our results cannot confirm in general the reliability of AI methods for this clinical condition due to small number of subjects with atypical absences considered in our study (2 cases). Moreover, sonified signals could be subjective, thus a validation by double-blind test should be evaluated in order to confirm the reliability of the discrimination of seizures events from false positives made by sonification [Bonebright, 2011]. Finally, the proposed sonification algorithm did not allow a recognition of false negative; another limit can be represented by the definition of only one possible reference signal (Section 1.3). All these aspects could be addressed in future developments of the methods in order to improve the performance of the absence seizure detectors both in offline and online mode.

In this work the possibility of finding a compromise between computational speed and accuracy for a fast and reliable identification of nonmotor generalized onset seizures was explored. To this aim, a specific filtering technique was proposed for the clinical problem under examination, compatible with its online implementation. In addition, an appropriate alert technique to inform the clinician about seizure onset was developed that requires low latency time. The results, though preliminary, suggest a possible optimization of existing methods applied to specific clinical settings. In this framework, absences may be considered as an 'optimal model' to test a seizure detection methodology, being aware that owing to their electroclinical characteristics, they are easily recognised and processed for sonification. However, the good results obtained, both in terms of balanced accuracy (about 96%) and latency times (1.25 s on average), offer the perspective for a real-time application of the methodology in the clinical setting of other types of seizures, which are more challenging for their automatic recognition and online sonification. The proposed methodology might also be helpful during video-EEG monitoring of candidates to epilepsy surgery, which requires a rapid intervention of the clinical staff at seizure onset and the real time evaluation of the related clinical signs, in order to better identify the seizure onset zone. In addition, our method might support clinicians in the interpretation of ictal video-EEG recordings through automatic event recognition techniques and the reduction from 19 simultaneous paths to a single sound information, that could be made through a predefined range of sounds combined with each characteristic event found in EEG data (e.g., seizure or interictal epileptiform discharges).

Moreover, the last part of the discussion opens to a new question:

Are there other frameworks where an online seizure detector could be a reliable and useful support to the clinical staff?

Neonatal Intensive Care Units (NICUs) could be among them, where a timely intervention and fast recognition of cerebral disorders (e.g. neonatal seizures) are utmost importance for the health of the newborns.

In conclusion, this chapter on absence seizure detection provides the motivations and methodological basis for the most substantial part of this doctoral thesis work: the development and evaluation of Neonatal Seizure Detectors for NICU that will be described in the next Chapters.

2. Automatic detection of seizures in Neonatal Intensive Care Units

Some contents in this chapter are based on the following publications:

- *Olmi, B., Frassineti, L., Lanatà, A. and Manfredi, C., 2021. Automatic Detection of Epileptic Seizures in Neonatal Intensive Care Units Through EEG, ECG and Video Recordings: A Survey. IEEE Access, vol. 9, pp. 138174-138191, 2021, doi: 10.1109/ACCESS.2021.3118227.*
- *Olmi, B., Manfredi, C., Frassineti, L., Dani, C., Lori, S., Bertini, G., Cossu, C., Bastianelli, M., Gabbanini, S., Lanatà, A., 2022. Heart Rate Variability Analysis for Seizure Detection in Neonatal Intensive Care Units. Bioengineering, Vol. 9, Issue 4, p. 165. https://doi.org/10.3390/bioengineering9040165*
- *Frassineti, L., Ermini, D., Fabbri, R., and Manfredi, C., 2020. Neonatal Seizures Detection using Stationary Wavelet Transform and Deep Neural Networks: Preliminary Results. In 20th IEEE Mediterranean Electrotechnical Conference, June 16-18, 2020, Palermo. doi: 10.1109/MELECON48756.2020.9140713.*

2.1 Introducing Neonatal Seizures

As discussed in Chapter 1 seizures and epilepsy may have different clinical manifestations, electroclinical characteristics and aetiologies, as well as varying incidence according to age. However, in the seizure classification proposed in [Fisher et al., 2017] (Figure 1.1) there is a missing piece regarding a specific kind of seizures: the neonatal seizures. In fact, ILAE in 2021 proposed a revision of classification and definition for seizures in neonates, that deserve a dedicated categorization and clinical evaluation [Pressler et al., 2021].

Neonatal seizures are the most common neurological emergency in the first days of life. The incidence is almost 1-5 per 1000 live births and 8.6/1000 in Neonatal Intensive Care Units (NICUs). Due to the specific category of patients, they do not fully match the criteria for diagnosis of epilepsy proposed in [Fisher et al., 2017]. Although the majority of neonatal seizures occur in the context of an acute illness, in some cases they could be a first manifestation of early infantile epilepsies [Pressler et al., 2021].

Since 1950, several works have proposed classification for neonatal seizures, according to clinical manifestations, autonomic nervous system changes [Volpe, 1989] or electroclinical characteristic [Mizrahi and Kellaway, 1987]. The American Clinical Neurophysiology Society defined an electrographic seizure as a "*a sudden, abnormal EEG event, defined by a repetitive and evolving patter with a minimum 2μV peak-to-peak voltage and duration of at least 10 seconds*" [Tsuchida et al., 2013]. This definition is partially in contrast with some electroclinical seizures in newborn that could have duration shorter than 10 seconds (e.g. myoclonic seizures), or the case of short rhythmic discharges (BIRDs), with duration always less than 10 seconds, which are suggest to be "very brief" electrographic seizures [Pressler et al., 2021].

Nowadays, as in [Pressler et al., 2021], such classifications are reconsidered, in order to uniform and standardize them, taking into account that neonatal seizures usually are characterized by a poor motor component (also known as subtle seizures or electrographic-only). In fact, at least the 50-80% of neonatal seizures are electrographic-only [Pressler and Lagae, 2020]. Moreover, there are several aetiologies behind neonatal seizures. As shown in Figure 2.1, the most common is the Hypoxic-Ischemic Encephalopathy (HIE) [Pressler et al., 2021], but there are also genetic, metabolic, or

cardiovascular origins, and some with unknown origin. Another point that must be considered is that the newborn's brain is constantly under development: immature neural connections, electroclinical characteristics as slower brain activities, degree of Autonomic Nervous System's development, make the neonatal seizure semiology different as compared to children and adults. The immature brain is characterized by high hyper-excitability due to poor inhibitory mechanisms and a surplus of excitatory neurotransmitters [Shellhaas, 2019]. Also, the effect of the preterm birth in the process of neurodevelopment could add another variability factor. Moreover, neonatal seizures are mainly focal, while only rarely they show a bilaterally distributed network.

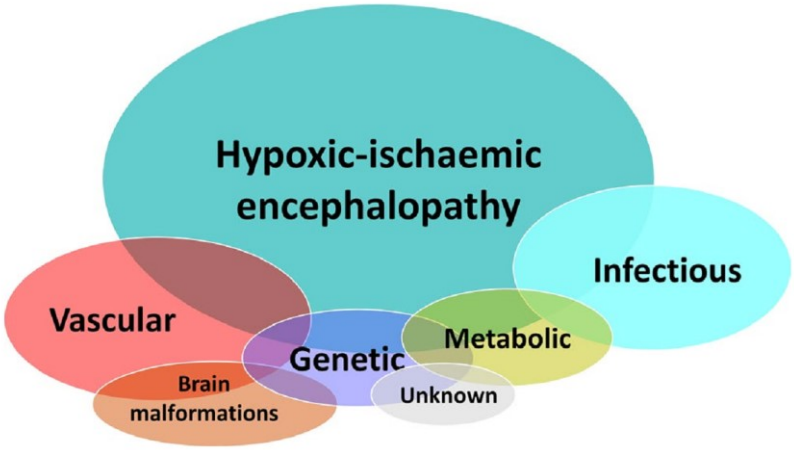


Figure 2.1. Common aetiologies of neonatal seizures, figure from [Pressler et al., 2021].

Here only a subset of possible differences between neonatal seizures and seizure in the child and in the adult are presented, but they are however sufficient to justify a dedicated definition and classification. Considering all these aspects the ILAE in [Pressler et al., 2021] proposed a novel conceptual definition:

"*An electrographic event with a pattern characterized by sudden, repetitive, evolving stereotyped waveforms with a beginning and end. The duration is not defined but has to be sufficient to demonstrate evolution in frequency and morphology of the discharges and needs to be long enough to allow recognition of onset, evolution, and resolution of an abnormal discharge.*"

Another important indication given in [Pressler et al., 2021] is the recognition of video-EEG as the gold-standard for diagnosis and characterization of neonatal seizures. Alternatives to multichannel EEG are the use of amplitude EEG (aEEG) or Color Spectral Density Array (CDSA), although such techniques have shown poor performance in terms of sensitivity for neonatal seizure detection [Shellhaas et al., 2007]. Anyway, aEEG and CDSA perform better than the seizure detection given by just the inspection of the clinical manifestations (e.g. by qualitative video analysis) [Malone et al., 2009]. Unfortunately, in several neonatal settings the EEG techniques and the expert staff for their interpretation are not available around the clock or not at all [Temko et al., 2017]. Moreover, even for expert staff the detection of neonatal seizures in video-EEG recordings is a non trivial and time-consuming task, with a high false negative rate that produce an inaccurate estimation of the seizure burden in the infants [Pavel et al., 2020, Frassineti et al., 2021a].

Since neonatal seizures have several origins, this implies that treatment depends strictly on their origin. In this PhD thesis a comprehensive discussion about the possible treatments of neonatal seizures is not provided. However, the most common treatment consists of anti-epileptic drugs (AEDs) and hypothermia [van Rooij et al., 2013]. It is noteworthy that newborns with seizures show

a remarkable resistance to pharmacological treatments, making the selection of the best treatment still an open issue in neonatology and paediatric neurology [Ramantani et al., 2019]. Moreover, it is important to remark that a missing diagnosis and recognition of seizure and its aetiology could have a negative impact on the neurodevelopment of the infants [Thibeault-Eybalin et al., 2009]. Thus, a fast and reliable detection of seizure events, the estimation of the seizure burden and monitoring of treatment's efficacy are utmost importance, especially in the context of NICUs, where timely decision is needed even when expert neurological staff is not available.

For these reasons, in the last years, there has been a growing interest in the development of Neonatal Seizure Detectors (NSDs) [Olmi et al., 2021, Temko and Lightbody 2016]. NSD could be defined as AI systems able to detect the time occurrence of neonatal seizures in online or offline mode, providing a support to the clinical staff for a timely detection and characterization of the seizure events. This can be obtained with the analysis of EEG signals or using all the sources of information, mainly physiological, available in neonatal settings such as NICUs.

This chapter aims at providing a survey of the current knowledge regarding NSDs as well as describing the development of an EEG-based NSD using data from a public dataset.

The chapter is organized as follows: in section 2.2 the state-of-the-art for NSDs is provided; section 2.3 shows an EEG-based NSD based on deep-learning algorithms, along with the description of the two neonatal seizure datasets used during this PhD study. In section 2.4 discussions on possible applications of NSDs in clinical practice and future perspectives are drawn, in order to introduce the next chapters of this PhD thesis.

2.2 State-of-the-art of Neonatal Seizure Detection

This section aims at providing a short overview about the state-of-the-art for NSDs, their limits and pitfalls, as well as their performance assessment. The search of suitable papers was performed in June 2021 based on the Scopus database using the following keywords: 'Neonatal seizure detection', that provided 1196 works. Then the research was refined considering only paper published in the last ten years and using the following MeSH terms: 'Automated systems / EEG monitoring / HRV / motion detection' AND 'Neonatal seizure', 'Seizure detection' AND 'NICU', 'image/video' AND 'processing' AND 'Neonatal seizure / NICU'. Papers based on single channel EEG, aEEG or CDSA were excluded. Then, the search was focused on those papers that proposed expert systems for automatic analysis of multi-channel EEG, ECG and video signals. Finally, 22 works were worth for the present analysis and they will be summarized in the next section.

2.2.1 Performance Assessment of NSDs

From the survey it clearly came out that there is a non-unanimous consensus about which metrics should be used in order to assess the performance of NSD methods. In other words, a standardized performance assessment framework for NSD is missing [Temko et al., 2011a].

The main metrics used in literature can be divided into three different categories: epoch-based metrics, event-based metrics and patient-based metrics.

Some of the epoch-based metrics were already introduced in Section 1.3.3. They are based on the segmentation of signals in time windows called "epochs". For a binary supervised absence seizure

classifier, the prediction could be divided into a positive class (usually the seizure epochs) and a negative class (interictal or seizure-free epochs). Conventionally the seizure detection problem was considered as binary, but these concepts could be easily extended to multiclass problems that could e.g. address the discrimination of different periods in the recordings. For example, the pre-ictal, interictal, post-ictal and ictal epoch detection is a four classes classification problem.

Thus, in this chapter only the binary formulation for the performance assessment is discussed. Metrics such as Accuracy (ACC), Sensitivity (SEN), False Positive Rate (FPR), $F1_{score}$, Matthews Correlation Coefficient (MCC) and Specificity (SPE) were already introduced in Chapter 1. As for absence seizures, they were widely used for the performance assessment of NSD. Most papers also reported the Receiver Operator Characteristic (ROC) curves plotting SEN vs SPE or SEN vs 1-SPE, or Precision vs Recall. Usually, from the ROC curves the Area Under the ROC Curve (AUC) is computed to compare the performance of different systems. Furthermore, when control patients (i.e. subjects with no seizures) are included in the evaluation also the AUC concatenated (AUC_{cc}) can be provided [Tapani et al., 2019], considering at the same time all the epochs from all the subjects to generate the ROC curve.

Another important category of metrics are the event-based metrics [Temko et al., 2011] that take into account the onset and offset of each seizure event, evaluating how the epochs were classified by the NSD. In other words, the time interval between the starting and ending time instant of seizure is defined as the "event". Although there is still no unanimous consensus regarding the exact definition of onset and offset for neonatal seizures [Frassineti et al., 2021a, Schindler et al., 2006] such metrics could be influenced by these definitions. Usually before computing such metrics post-processing methods were added, such as "collar techniques", or spatial threshold if a multi-channel EEG was used [Temko et al. 2017].

The main event-based metrics are:

- Good Detection Rate (GDR): the overall percentage of seizure events correctly identified by the NSD [Temko et al. 2011a].
- False Discovery Rate (FDR): the overall percentage of seizure events incorrectly identified by the NSD.
- False Detection per Hour (FDH or FD/h): describes the number of false seizure events identified by the NSD that have no overlap with the previous event labelled by the expert [Temko et al. 2011a].
- Mean False Detection Duration (MFDD): the average duration of all false detections without any overlap among them [Temko et al. 2011a].

Regarding the patient-based metrics they strongly depend on the context and the framework, however most popular are: the sensitivity and specificity of NSD to detect newborn with seizures (a sort of index of seizure risk [Frassineti et al., 2021b, Pavel et al. , 2020]) and the estimation of the seizure burden (i.e. the number of seizure events detected by the NSD) [Pavel et al. , 2020].

Moreover, a crucial point during the development of an NSD is the method adopted for its validation. Popular validation methods such as k-fold-cross validation or hold-out validation can be used in the context of NSD, however since often the number of patients is low and machine-learning methods need a huge amount of data, these approaches tend to overestimate the performance of the NSD, thus living rise to a sort of patient-specific NSD [Temko et al. 2011a]. Instead, the goal of an ideal NSD is to be a reliable patient-independent detector [Temko et al. 2017]. In order to achieve this result for limited dataset, the Leave-One-Subject-Out (LOSO) validation should be preferred [Olmi et al. 2021,

Temko et al. 2011a, Gotman et al., 1997]. Basically, if a huge amount of data would be available, even k-fold-, hold-out- cross validation and other methods could be used. However, to date, the largest public dataset for neonatal seizures consists of 79 subjects [Stevenson et al., 2019], that is not comparable to other datasets used for AI research [Hastie et al., 2001]. This introduces one of the main current pitfalls for AI methods for NSD, that is the limited amount of available datasets. This aspect will be further discussed in Section 2.3.

2.2.2 Heuristic algorithms for EEG-based NSD

The first attempt in literature for the development of NSD was based on heuristic algorithms (HA). HA are based on the definition of rules, threshold values and specific parameters obtained testing the data and according to pathophysiological knowledge of neonatal seizures. Usually, HA combined the analysis of the morphology of EEG signals, searching for variations of the background activity on analogy to the perceptual evaluation made by the neurologist, such as looking for repetitive and specific waveforms. The main papers based on HA are shown in Table 2.1. All the reported papers used private datasets.

Table 2.1. Main NSD systems based on EEG and HA. [Olmi et al., 2021].

Paper	Method	Dataset (newborns)	Epoch length	Performance
Liu et al. 1992	Autocorrelation analysis	14	30s	SEN=84% SPE=98%
Gotman et al. 1997	Detection of three characteristics patterns	55	10s, 75% overlap	SEN=71%
Deburchgraeve et al. 2008	Detection of two/four major seizure patterns	26	5s, 4s overlap	SEN=88% FDH=0.66 h^{-1}
Navakatikyan et al. 2006	Detection of increased regularity in EEG wave sequences	66	N.A.	SEN= 83-95%

2.2.3 Data-driven algorithms for EEG-based NSD

Data-driven algorithms for NSDs are the most investigated in literature [Olmi et al. 2021]. The absence seizure detector proposed in [Frassineti et al., 2019] can be considered part of the data-driven algorithms. These approaches use machine-learning (ML) techniques based on the extraction of features to characterize the data. As already explained in Section 1.3.3, supervised ML models are built by the process of training and validation using features and expert's labels. Generally, features are extracted from EEG epochs, with or without overlap. Features can be grouped in the frequency, time and information theory domains.

As shown in Table 2.2 the most popular AI model is the Support Vector Machine (SVM) one. However, these approaches have recently come under criticism: they require an extensive evaluation of which subset of hand-crafted features could be used, and they may not be optimal in some framework [O'Shea et al., 2020]. This limit led to the investigation of other methods such as deep-learning algorithms that will be introduced in subsection 2.2.4. Nevertheless, data-driven algorithms are still widely used, also because most of them make explicit the process employed to discriminate between seizure epochs and non-seizure epochs [Olmi et al., 2021]. Indeed, the work by Pavel et al. [Pavel et al., 2020] is one of the first multi-centre clinical validation for an EEG-based NSD.

Moreover Pavel et al. proved that NSD given a more reliable estimation of the seizure burden in newborns, with shorter time for diagnosis and the evaluation of efficacy of the treatment adopted, confirming the potentiality of EEG-based NSD in clinical practice. Finally, several works used mainly single-channel EEG features, thus a possible improvement of such methods might be the use of multi-channel EEG features that for seizure detection in adults have been proved to better characterize the dynamics during seizure events [Frassineti et al., 2021a]. The main works who used data-driven algorithms for EEG-based NSDs are reported in Table 2.2.

Table 2.2. Main EEG-based NSDs using data-driven algorithms. [Olmi et al., 2021].

Paper	Method	Dataset (newborns)	Validation	Epoch length	Performance
Thomas et al. 2010	GMM	55 with HIE	LOSO	8s, 50% overlap	GDR=79% FDH= 0.5 h^{-1} MFDD= 2 min SPE=93% SEN=76%
Temko et al. 2011b	SVM	55 with HIE	LOSO	8s, 50% overlap	GDR=100% FDH= 4 h^{-1}
Pavel et al. 2020	ANSeR	258	N.A.	N.A.	Patient-based metrics Non-algorithm group: SEN=89%, SPE=89%, FDR=22% Algorithm group: SEN=81%, SPE=84%, FDR=36% Recognition of seizure hours Non-algorithm group SEN=45%, Algorithm group SEN=66%
Tapani et al. 2019	SVM & autocorrelation analysis	79	LOSO	32s, 28s overlap	AUC=92% SEN=76% SPE=99%

2.2.4 Deep-Learning algorithms for EEG-based NSD

For data-driven algorithms the choice of the features is a critical point that determines the NSD's performance [Olmi et al., 2021, O'Shea et al., 2020]. To overcome such pitfall several deep-learning (DL) algorithms were proposed as EEG-based NSD. DL methods do not require hand-crafted features as in data-driven algorithms. In the literature several DL architectures were analysed and tested, such as Convolutional Neural Network (CNN) and Fully Convolutional Neural Networks (FCN), using as input both the original EEG epochs or a time-frequency representation of them. Usually time-frequency representations are obtained by the Short Time Fourier Transform (STFT), the Wavelet Transform (WT) or other source decomposition methods. The time-frequency transformation allows to treat the multichannel EEG signals as a sort of image as input for the DL methods. Also Transfer Learning methods were evaluated to develop DL EEG-based NSD [Caliskan and Rencuzogullari, 2021]. Though preliminary, results obtained with DL methods are emerging as the most promising approach in the NSD field, and they are going to be proposed for clinical validation. Some DL methods showed high performance both on term and preterm newborns [O'Shea et al., 2021].

However deep-learning algorithms are still considered as a sort of *black-box* models for NSD and further studies are needed in order to explain the decision rules that such algorithms apply for classification [Olmi et al., 2021]. Another remarkable limit for some of the DL methods, except for pre-trained deep-neural networks, is that they may require a computational time often not suitable for real time applications, allowing only off-line analysis. The main works that used deep-learning algorithms for EEG-based NSDs are reported in Table 2.3.

Table 2.3. Main EEG-based NSDs using deep-learning algorithms. [Olmi et al., 2021].

Paper	Method	Dataset (newborns)	Validation	Epoch length	Performance
Ansari et al. 2019	CNN + Random Forest	48 with HIE	Hold out	90s, 60s overlap	AUC=83%, SEN=83%, SPE=78%, GDR=77%, FDH=0.90 h^{-1}
O' Shea et al. 2020	2D FCN	55 with HIE + 79	LOSO	8s,	AUC_{cc}=95.6%
Tanveer et al. 2021	2D CNN	39	10-fold cross validation	1s, 50% overlap	ACC= 96%, AUC=99.3%
Caliskan et al. 2021	p-CNN	39	Hold-out	30s, 2s shift	Patient specific AUC=99%
O' Shea et al. 2021	FCN	16 preterms	N.A.	8s, overlap 4s and 7s	AUC=95%

2.2.5 ECG-based and Video-based NSD

All the methods described in subsections 2.2.2 - 2.2.4 have in common the use of EEG as source of information for NSD. However, as already noticed in sect. 2.1, such signal could not be available in some neonatal settings. Thus, in the literature there was a growing interest in NSDs that used or combined source of information different from EEG. Most of the works found in the present survey used Electrocardiography (ECG) signals, in particular Heart Rate Variability (HRV) and video signals. ECG and video are considered a more accessible source, simple, non-invasive and less expensive than EEG [Olmi et al., 2021]. How and why HRV analysis could be helpful to detect and characterize neonatal seizures will be discussed in detail in Chapter 4.

By analogy to EEG-based methods, ECG- and HRV-based ones can be divided in heuristic, data-driven and deep-learning algorithms. The main difference lies in the epoch length, as ECG epochs are generally longer than in EEG because the dynamics involved are quite slower than the EEG ones. This introduces one of the main limitations of ECG-based NSD: at present such methods cannot be used for on-line applications [Frassineti et al., 2021b, Frassineti et al., 2021c]. Based on the promising results obtained on ECG-based NSD, in the literature the combination of EEG and ECG features for data-driven algorithms was proposed. [Greene et al., 2007a, Mesbah et al., 2012], proved that the multimodal analysis gave higher performance than the use of only EEG or ECG/HRV features. The main works regarding ECG-based NSDs and NSDs that used combination of features from ECG and EEG are reported in Table 2.4.

Table 2.4. Main ECG-based NSDs and NSDs that used both EEG and ECG signals. [Olmi et al., 2021].

Paper	Method	Dataset (newborns)	Validation	Epoch length	Performance
Malarvili et al. 2009	Heuristic TFD	5	LOSO	64s	SEN=83% SPE=100%
Greene et al. 2007b	LD	7	LOSO	60s	ACC=61% SEN=78% SPE=51%
Doyle et al. 2010	SVM	55 with HIE	LOSO	60s	AUC=60% SEN=60%
Greene et al. 2007a	LD EEG+ECG features	10	LOSO	60s	GDR=81% FDR=33% ACC=68% SEN=74% SPE=66%
Mesbah et al. 2012	1-NN + linear classifiers (EEG+ECG features)	8	LOSO	64s	SEN=95% SPE=94%
Temko et al. 2015b	SVM, EEG+ECG features	38	LOSO	60s	AUC=86%

Regarding video analysis, it has been proven that even in subtle seizures little movements of face, arms and legs are present [Volpe, 1989, Mizrahi and Kellaway, 1987]. Indeed, it is well known that the clinical manifestations of seizures (e.g. abnormal movements in clonic seizures) can be easily detected by automatic video analysis [Malone et al., 2009]. Thus, a detailed analysis of AI algorithms applied on video recordings could show details that are hard to detect by visual inspection [Olmi et al., 2021]. In fact, as shown in Table 2.5, such video features were already used in several works, suggesting that video analysis could be considered as a valid NSD or in combination with other physiological signals such as EEG or ECG. The main papers that used video analysis for NSD are reported in Table 2.5.

Table 2.5. Main video based NSDs. [Olmi et al., 2021].

Paper	Method	Dataset (newborns)	Validation	Epoch length	Performance	Notes
Pisani et al. 2014	Frame Differencing	12	Binary Statistical Analysis	10s, 50% overlap	AUC=79% SEN=71% SPE=69%	Distinction between clonic seizures and random movements
Cattani et al. 2017	Frame Differencing (3 cameras)	1	N.A.	10s, 50% overlap	SEN=90% SPE=90%	Detection of clonic seizures
Karayiannis et al. 2005	FFNN, optical flow analysis	43	Hold-out	N.A.	Test set SEN < 90% SPE > 90%	Distiction between mycolonic and focal clonic seizures

The papers summarized in subsections 2.2.3 - 2.2.5 did not provide information about aetiologies of the detected seizures. However, some of them (e.g. [Deburchgraeve et al. 2008]) pointed out that the cause of seizure events could be identified by analyzing seizure events themselves. Therefore, NSD systems able to automatically characterize the aetiologies investigating the available electrophysiological and clinical signals could provide an additional support to clinicians. Indeed, identifying aetiologies is crucial to determine specific pharmacological treatments and subsequent prognoses [Pressler et al., 2021].

2.3 Neonatal Seizures Datasets: the lack of public data

As already said in previous sections one of the main limits in the field of neonatal seizure detection is the availability of public datasets. Several papers presented in the survey used private datasets. This implies that the comparison of the existing systems is challenging, as well as the already mentioned non-uniformity regarding the metrics used for the performance assessment of NSDs.

Until 2022 the only public dataset available containing neonatal EEG and ECG recordings is the Helsinki Dataset (hereinafter denoted as HD) [Stevenson et al., 2019]. It collects multi-channel EEG signals from 79 full-term newborns ad the NICU of the Helsinki University Hospital. The recordings have a mean duration of 1h, using 19 electrodes for EEG in the so-called double-banana montage. The sampling frequency was 256 Hz both for EEG and ECG. The main seizure's aetiology was HIE that concerned at least 50% of the newborns with seizures. The dataset was labelled by three experts separately. As investigated by Tapani et al. [Tapani et al., 2019], the agreement across the experts was not perfect, confirming that the neonatal seizure detection is not a trivial task even for experienced clinical staff. In fact, among the 79 subjects 39 have seizure events with unanimous consensus of the three experts, 22 are seizure-free with unanimous consensus, while the remaining 18 obtained different labels from the experts. Moreover, the experts identify different time occurrence of onset and offset for several seizure events. This opens another important field of research for NSD: evaluating how expert's labels could vary the NSD's performance. However, it is important to remark that the labelling process is highly time-consuming, thus in literature it is much more likely to found labels from a single expert.

During the PhD period several of the methodologies presented in this thesis were initially implemented and evaluated using data from the HD [Frassineti et al. 2021a, Frassineti et al., 2021b, Frassineti et al., 2021c, Frassineti et al., 2021d, Frassineti et al., 2020]. To address the lack of data, during the PhD period a proper private data set was collected, that was used to test the reproducibility of the AI methods developed using the HD dataset.

The new dataset of video-EEG and ECG recording was collected at the Neuro-physiopathology and Neonatology Clinical Units of AOU Careggi, Firenze, Italy. Data were retrospectively collected between March 2010 and October 2020. The dataset consists of 51 full-term newborns with gestational age between 38 and 41 weeks. Out of 51 subjects, 29 were control patients without any seizure events. The remaining 22 had seizure events in their recordings, labelled by an expert neurologist of the AOU Careggi clinical staff. Ten out of the 22 newborns with seizures showed electrographic-only seizures (EGP), the remaining 12 exhibited electroclinical seizures (ECP). As for HD, the main aetiology was the HIE (12/22 subjects). None of the considered newborns have heart disease. The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of Careggi University Hospital, Firenze, Italy. The Local Ethical Committee approval code is 02/2013. EEG and ECG were synchronously recorded using Nemus ICU

Galileo NT Line system (EB Neuro S.p.A., Firenze, Italy). The sampling frequency was 128 Hz. For ECG signals, the one-lead ECG system was equipped with two electrodes: the active one was placed on the left axillary line at the level of the 7th rib, while the reference electrodes was placed on the right clavicle. The mean time duration of recordings for each patient was 53 minutes, the overall duration of the dataset was about 45 hours. The mean seizure duration per patient was 00:09:39 h. Other information regarding the newborns with seizures are summarized on Table 2.6 and in [Olmi et al., 2022a].

Video recordings were available only for 42 subjects, of which 22 with seizure events. The total length of video recordings was 39:59:08 h (mean duration per patient 00:57:07 h). The videos were collected using the Panasonic's Super-Dynamic Colour Camera WV-CP450/G. Sampling rate: 25 frame/s. Resolution: 640x480 pixels. As for HD, the Careggi Dataset (hereinafter named CD) was used to implement and validate various methods presented in this thesis [Olmi et al., 2022a, Olmi et al., 2022b] and they will be discussed in Chapter 4 and 5.

Table 2.6. Description of recordings of patients with electrographic (EGP) and electroclinical (ECP) seizures. [Olmi et al., 2021].

Patients	Record Length (h)	Number of Seizure Events	Seizure Events Duration			Etiology
			Average	Min	Max	
EGP1	01:08:29	1	00:03:40	00:03:40	00:03:40	Metabolic
EGP2	01:09:34	8	00:00:24	00:00:12	00:00:43	HIE
EGP3	01:10:15	1	00:00:31	00:00:31	00:00:31	Other
EGP4	00:56:28	4	00:00:16	00:00:07	00:00:22	HIE
EGP5	01:02:30	3	00:01:45	00:01:18	00:02:00	HIE
EGP6	01:17:02	4	00:00:12	00:00:08	00:00:16	HIE
EGP7	01:05:53	1	00:00:50	00:00:50	00:00:50	HIE
EGP8	00:45:40	3	00:01:28	00:01:19	00:01:40	HIE
EGP9	01:20:39	1	00:00:29	00:00:29	00:00:29	Other
EGP10	00:58:22	1	00:01:40	00:01:40	00:01:40	Other
ECP1	00:48:12	1	00:48:12	00:48:12	00:48:12	Genetic
ECP2	00:54:31	13	00:01:48	00:00:27	00:09:49	Metabolic
ECP3	01:04:14	6	00:02:49	00:01:20	00:04:40	Stroke
ECP4	01:18:36	5	00:00:33	00:00:16	00:00:50	Genetic
ECP5	01:00:10	3	00:07:54	00:01:56	00:19:45	Other
ECP6	01:10:02	7	00:01:05	00:00:33	00:02:23	HIE
ECP7	00:54:06	1	00:10:26	00:10:26	00:10:26	Stroke
ECP8	01:00:43	10	00:02:53	00:01:19	00:11:04	Other
ECP9	01:30:08	5	00:00:48	00:00:22	00:01:16	HIE
ECP10	00:41:16	3	00:03:52	00:00:20	00:10:10	Stroke
ECP11	01:01:02	8	00:00:52	00:00:26	00:02:01	HIE
ECP12	01:20:06	3	00:01:58	00:01:19	00:02:31	HIE
Total	23:37:58	92				

2.4. Development of a DL EEG-based NSD: proof of concept and preliminary results

2.4.1 Introduction

Some contents in this section are based on the following publication:

- *Frassineti, L., Ermini, D., Fabbri, R., Manfredi, C., 2020. Neonatal Seizures Detection using Stationary Wavelet Transform and Deep Neural Networks: Preliminary Results. 2020 IEEE 20th Mediterranean Electrotechnical Conference (MELECON). https://doi.org/10.1109/melecon48756.2020.9140713*

In this section a Deep-Learning (DL) EEG-based NSD is provided that takes into account the main drawbacks of data-driven algorithms for NSD, the use of hand-crafted features and their possible non optimal behaviour in some cases. To overcome such problems, an hybrid techniques was developed that combines SWT applied to raw EEG as input to Convolutional Neural Networks (CNN) and Fully Convolutional Neural Network (FCN). The method was developed on analogy to the hybrid techniques proposed by Ansari et al. [Ansari et al., 2019] that applies CNN and Random Forest. The use of FCN has already proven useful for NSD by O'Shea et al. [O'Shea et al., 2021, O'Shea et al., 2020].

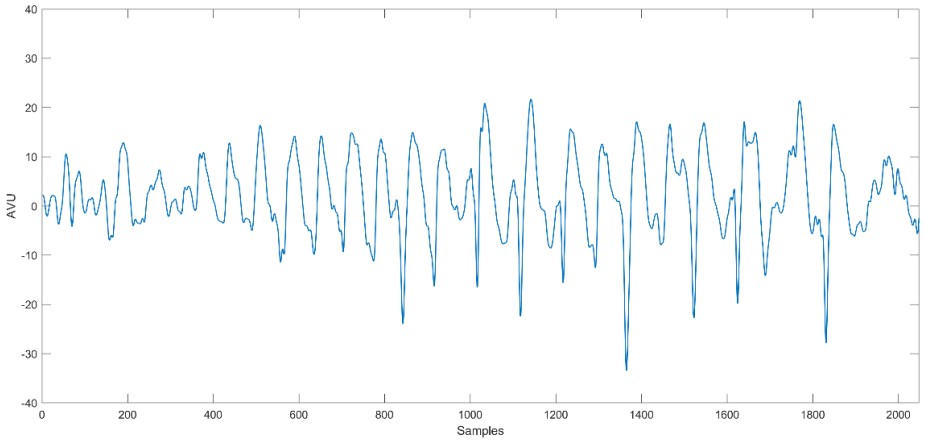
The SWT was added to remove unnecessary frequency information for the NSD problem and perform a sort of data-augmentation on the EEG signals [Shorten and Khoshgoftaar, 2019] representing each derivation though the details and approximation levels, on analogy to EEG time-frequency representation already used as input images to neural networks [Roy et al., 2019]. All the methods were trained and tested on the Helsinki dataset (HD) introduced in Section 2.3. Only the newborns who received unanimous consensus among experts were considered. Thus, the analysis was performed on 39 newborns with seizures and 22 seizure-free. The AND operator was used to merge the expert labels [Frassineti et al., 2020]. Until 2020 this was the first attempt to use SWT+CNN/FCN both in training and validation with the HD. O'Shea et al. used the HD but only as test set [O'Shea et al., 2020].

The proposed methods are implemented under MATLAB computing environment (Deep Learning Toolbox, version 2019b), OS Windows 10, 64 bit. Processor: AMD Ryzen 5 2600 Six-Core, CPU 3.40 GHz, RAM 16 Gb, GPU NVIDIA GeForce GT 1030. We remark that in this work only the HD dataset was considered, while analysis and evaluations on CD dataset will be discussed in Chapter 4 and 5.

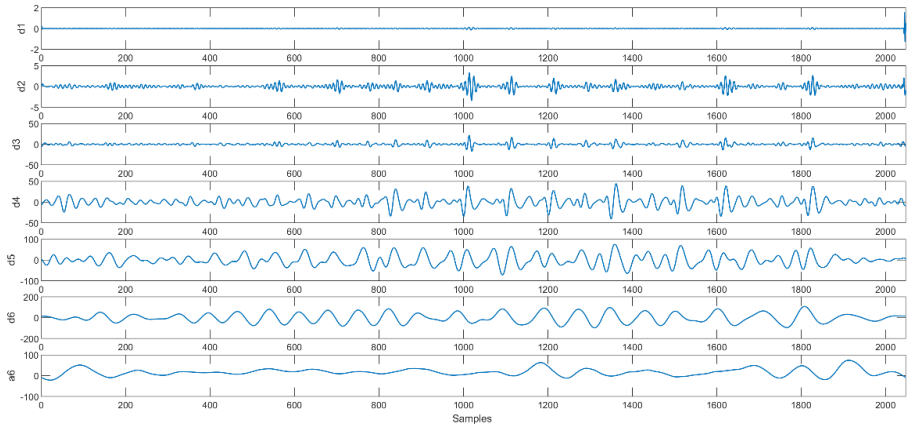
2.4.2 Material and Methods

In this work the same montage proposed in [O'Shea et al., 2020] was used: F4-C4, C4-O2, F3-C3, C3-O1, T4-C4, C4-Cz, Cz-C3, C3-T3. Three different epoch lengths were investigated: 4s, 8s and 16s with 50% overlap. Different lengths were considered as there is still no unanimous consensus about which could be the best time windows for neonatal seizure detection [Olmi et al, 2021]. Each window was filtered with a band-pass filter (0.5-32 Hz) and Notch filter (50Hz). For SWT the Daubechies 5 (db5) mother wavelet function was used [Ekim et al., 2017]. Figure 2.2 shows an example of SWT decomposition for a 8s epoch (only one EEG channel is displayed). Using six levels of decomposition, each EEG epoch of size 1xN, where N is number of samples in the epoch, was expanded into a 7xN matrix, where each row represents the corresponding level of details and

approximation. According to [Baboukani et al., 2019], for each derivation the details level d6, d5, d4 and the approximation level a6 were taken into account. This choice was made as they approximately correspond to the conventional EEG rhythms for a sampling frequency of 256Hz. Thus, the other decomposition levels were removed, obtaining a 4xN matrix. Therefore, for an 8s window with sampling frequency 256 Hz, a 4x2048 matrix is obtained for each derivation. Then the eight EEG derivations in SWT representations were concatenated, generating a 32x2048 matrix. Finally, a down-sampling step was added to have windows with 256 samples for each row. This 32x256 matrix was used as input for the proposed neural networks. Considering only the consensus seizures, they were 39259 seizures labels and 363566 non-seizures labels, where a label represents the corresponding time instant (in seconds) of the EEG exam (at 256 Hz sampling frequency). Thus, for example considering the 8-seconds windows and the 50% overlap, we obtained 10239 seizures windows and 90342 non-seizures windows with an imbalance ratio of 8.82.



(a)



(b)

Figure 2.2 (a) Example of single channel EEG showing a seizure. (b) Its corresponding 6-levels SWT. Epoch length: 8s. [Frassineti et al. 2020]

In Table 2.7 the CNN and FCN architecture properties are shown. For all the cases studied (epochs 4s, 8s and 16s) both CCN and FCN were evaluated. A detailed description of the layers employed for the networks can be found in [Goodfellow et al., 2016]. The final parameters used in the training phase of network are the following: Solver stochastic gradient descent with momentum [Qian, 1999] with learning rate 0.001 and momentum 0.9, Mini-Batch Size 128, Validation Patience 8. Regarding the activation functions also other functions were preliminary investigated such as Rectified Linear

Unit (ReLu) and LeakyReLu [Mass et al., 2013], but in our case the hyperbolic tangent layer (Tanh) gave a better performance. The architectures and the parameters used in the training phase were obtained by a try-and-error process, thus they might not represent the optimal configuration. During the training of the networks, to address the imbalance between seizure epochs and seizure-free epochs, a random down-sampling for each patient was applied to obtain a ratio around 1:1. To avoid over-fitting of the networks, during the training phase an internal k-fold cross-validation was applied, using 5 patients randomly selected from the original training set. Then, the LOSO validation was used to quantify the performance of the DL methods. For each patient in the validation step all the epochs were used without any down-sampling as in the training step.

The following epoch-based metrics are applied [Temko et al., 2011a]: ACC, SEN, SPE, $F1_{score}$, MCC, AUC and AUC_{cc}. Moreover, the following event-based metrics are evaluated: GDR, FD/h and the latency time [Temko et al., 2017]. To choose the best DL methods in validation the parameter AUC_{cc} was considered. Results related to these validations will be presented in subsection 2.4.3.

Table 2.7. Architecture properties. [Frassineti et al., 2020].

Feature Extraction Blocks (CNN and FCN)	
Input Layer 32x256 rescale symmetric [-1,1] Convolution 2d Layer, filter size (2,4), 32 filters, Stride (2,2) Batch Normalization Layer Max Pooling Layer, pool size (1,2), Stride (1,2) Tanh Convolution 2d Layer filter size (2,4), 64 filters, Stride (2,2) Batch Normalization Layer Max Pooling Layer, pool size (1,2), Stride (1,2) Tanh Convolution 2d Layer filter size (1,4), 128 filters, Stride (1,1) Batch Normalization Layer Max Pooling Layer, pool size (1,2), Stride (1,2) Tanh	
Classification Blocks	
CNN	**FCN**
Fully Connected Layer (1024x1024) Dropout (0.5) Fully Connected Layer (2x1024) Sofmax Layer Classification Layer (Cross-Entropy)	Convolution 2d Layer, filter size (1,3), 2 filters, Stride (1,2) Average Pooling, pool size (1,2) Max Pooling, pool size (8,1) Softmax Layer Classification Layer (Cross-Entropy)

Finally, to assess if SWT operation could be helpful or redundant, two different tests (hereinafter Test1 and Test2) were performed. First another EEG matrix 8x256 was built for each epoch without using the SWT operation, thus keeping the original time series. The second test consisted in using an EEG matrix 32x256 but in this case each derivation was replicated for each detail and approximation level considered. This test aimed to establish if any increase in performance could be due to the application of SWT or only to the concatenation step.

2.4.3 Results

In table 2.8 and 2.9 the results of CNN and FCN using different epoch lengths are shown. Only the epoch-based metrics are shown. The 8s windows with FCN gave the best performance in terms of

AUC_{cc}. To assess redundancy of the SWT decomposition, Table 2.10 reports the comparison between the best case (FCN 8s) and the corresponding Test1 and Test2. For this case also the event-based metrics are reported. The ROC curves related to the parameters AUC_{cc} are shown in Figure 2.3. The results shown in Tables 2.8, 2.9 and 2.10 are related to the average values for all the metrics considered during LOSO validation. Because of the random selection of the patients in the internal cross-validation, it has been investigated if this approach could significantly alter the performance. To this aim, all the described procedures were repeated three times with the five patients used, without finding significant differences in the performances (less than 1-2% in Accuracy).

Table 2.8. Results LOSO validation using CNN (average values). [Frassineti et al., 2020].

Epoch Length	AUC (%)	AUC_{cc} (%)	ACC (%)	SEN (%)	SPE (%)	F1score (%)	MCC
4s	77	76	83	39	92	40	0.28
8s	77	79	79	51	81	41	0.27
16s	75	77	85	37	94	49	0.30

Table 2.9. Results LOSO validation using FCN (average values). [Frassineti et al., 2020].

Epoch Length	AUC (%)	AUC_{cc} (%)	ACC (%)	SEN (%)	SPE (%)	F1score (%)	MCC
4s	79	83	84	47	89	42	0.29
8s	81	87	82	63	83	48	0.35
16s	79	84	80	57	82	45	0.31

Table 2.10. Comparison between SWT-FCN (8s epoch length), Test1 and Test2. [Frassineti et al., 2020].

	Epoch-based metrics							Event-based metrics		
Test	AUC (%)	AUC_{cc} (%)	ACC (%)	SEN (%)	SPE (%)	F1score (%)	MCC	GDR (%)	FD/h (h-1)	Latency Time (s)
8s FCN	81	87	82	63	83	48	0.35	78	1.6	32
Test1	76	81	78	55	80	41	0.26	75	1.9	40
Test2	75	81	81	50	84	40	0.27	77	2	32

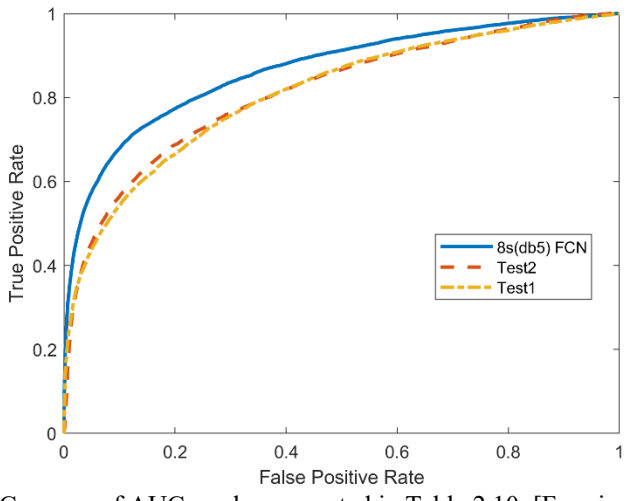


Figure 2.3 ROC curves of AUC_{cc} values reported in Table 2.10. [Frassineti et al., 2020].

2.4.4 Discussion and Conclusions

In this work a proof of concept about the combination of SWT and neural networks as NSD was provided. Though preliminary, the results confirm that the use of SWT could be helpful in increasing the performance of the seizure detectors. For DL EEG-based, SWT could be considered as sort of data-augmentation techniques. In fact, as shown in Table 2.10 AUC_{cc} for SWT+FCN was larger than for Test1 and Test2. Moreover, as shown in Table 2.8 and 2.9, FCN gave a higher performance with respect to CNN, confirming what already suggested by O'Shea et al. [O'Shea et al., 2020].

Furthermore, considering the results obtained using different epoch lengths, it was confirmed that the choice of the time duration of windows is a crucial and critical point in the development of any NSD. In our case the best epoch length was 8 seconds, as in [O'Shea et al., 2020]. The reason could be that, increasing the scale of observation, the useful details for a correct discrimination between seizure and not seizure become a negligible part of the available information. As shown in Table 2.10, the latency times obtained suggest that the systems can provide a reply to the user almost two times per minute (32, 40 seconds). This timing could provide an almost constant feedback about the brain condition of the newborn in NICU, and thus could be a valid support for an early detection of possible seizures events and timely application of the most suitable therapy for each patient.

The performance obtained for the 8s window are close to those presented in [O'Shea et al., 2020], both for epoch-based metrics (AUC 83%) and event-based metrics (GDR 77% and FD/h 0.90). However, they are still lower than those obtained by O'Shea et al. in [O'Shea et al., 2020]. One reason could be that the training dataset used here was smaller than the one in [O'Shea et al., 2020]. Indeed, the authors used recordings with an overall duration of 834h with 1389 seizure events. Instead, the overall duration of the recordings in the dataset used here amounts to 112h and there are only 342 seizure events. It is well-known that the deep-learning performance strictly depends on the available training data set; therefore, the possibility to have a large data set may represent one of the basic requirements for the reliability of the methods proposed in this field.

Future work will be devoted to the neural network architectures used and to the choice of the mother-wavelet function for the SWT. Furthermore, to enlarge the case studies and address the unbalancing problem between classes, the use of the Generative Adversarial Networks (GAN) may be explored [Goodfellow et al., 2020].

In conclusion, the work presented here is one of the first approaches to the application and testing of deep-learning methods on a public dataset of neonatal seizures. It provides first indications about the use of the Stationary Wavelet Transform in combination with deep-learning algorithms to solve the NSD problem, using a public dataset with multi-expert labels. The encouraging results show the possibility of successfully using these hybrid techniques. Moreover, they pave the way for a novel approach to perform data-augmentation for EEG time series.

3. Multiparametric characterization of neonatal seizures by EEG multichannel features

Some contents in this chapter are based on the following publication:

- *Frassineti, L., Parente, A., Manfredi, C., 2021. Multiparametric EEG analysis of brain network dynamics during neonatal seizures. Journal of Neuroscience Methods, Vol. 348, p. 109003. https://doi.org/10.1016/j.jneumeth.2020.109003*

In Chapter 2 the lack of information in the literature about multichannel EEG features applied to NSD has been highlighted. Most of the existing NSDs used single-channel EEG features, because in NICUs usually the EEG montages consist of a limited number of electrodes. However, it was proven that multichannel EEG techniques are able to detect more helpful information for neonatal seizures than montages with a limited number of electrodes, such as aEEG [Baboukani et al., 2019, Shellhaas et al., 2007]. Thus, although promising, the use of multi-channel EEG features for neonatal seizure detection was still not fully addressed in literature. Indeed, in adults and children such approaches have already shown their usefulness in seizure detection and characterization. Moreover, several EEG features used in NSDs did not provide information about the brain time-varying dynamics of the seizure, being based on static information of the analysed EEG signal to discriminate seizure periods from non-seizure periods [Frassineti et al., 2021a].

Therefore, in this chapter the use of multichannel EEG features in newborns will be discussed. Moreover, results of two different measures: Synchronizability [Schindler et al., 2008] and Circular Omega Complexity [Baboukani et al., 2019], for the characterization of pre-ictal, ictal and post-ictal periods will be described and discussed. The chapter is organized as follows: in section 3.1 a short background about multiparametric EEG analysis for seizure detection and characterization is provided; section 3.2 introduces the methods applied to the Helsinki Dataset [Stevenson et al., 2019] in order to characterize pre- post- and ictal periods in newborns. Results are shown in section 3.3., while discussion of the results and conclusions are drawn in section 3.4.

3.1 Background

In the last years, the study of brain network dynamics by EEG has been proposed in several clinical applications, such as: evaluating co-mordibities with the mechanisms of cognitive decline, predicting the risk of developing seizures [van Diessen et al., 2013]; finding biomarkers for brain tumours [Douw et al., 2010]; characterizing aetiologies in epilepsy [Kuchenbuch et al., 2019]; evaluating the brain state after surgical intervention [Goodfellow et al., 2016]. Moreover, several studies related to connectomics [Fornito et al., 2016] suggested that epilepsy could be a brain network disease [Lehnertz et al., 2017]. Indeed, the epileptic seizure focus, i.e. the area of the brain where the seizure starts, could be only a part of the complex dynamics of this syndrome. This allows overcoming the concept that seizures are not just hyper-synchronous brain states [Jiruska et al., 2013], but only one part of the complex mosaic behind ictal events, meaning that seizures are dynamical brain processes. Therefore, functional and structural analysis of brain networks may identify specific patterns inside the cerebral network related to specific epileptic syndromes.

A traditional approach for characterizing the dynamics of brain complex dynamics in the adults is the phase synchrony analysis (PSA) [Smith et al., 2019]. However, it has been proven that PSA methods such as the Phase Locking Value, may not be able to outline the global dynamics in multivariate dynamics. [Smith et al., 2019, Baboukani et al., 2019]. To overcome such limitation several multivariate and multiparametric approaches have been proposed, such as general field synchronization and generalized phase synchrony analysis [Omidvarnia et al., 2013], empirical mode decomposition [Mutlu et al., 2010], phase lag index [Stam et al., 2007], hyper-dimensional geometry [Al-Khassaweneh et al., 2016] and S-estimator measures [Oshima et al., 2006]. However, most of the findings in literature are mainly related to EEG analysis in adults or in children. For the reasons introduced on Chapter 2, a direct transposition of them in newborns cannot be done, or at least must be considered with caution. It is important to remark that electro-clinical characteristics of neonatal EEG are different from those in adults. Therefore, it is necessary to define and validate novel combinations of multivariate analysis to assess findings related to the neonatal seizure dynamics. In fact, only few studies investigated multivariate dynamics during neonatal seizures and in general in neonatal brain network dynamics. Preliminary analysis performed by Baboukani et al. [Baboukani et al., 2019] showed that neonatal seizures could be described by multivariate features and as input of NSD. The analysis of functional connectivity carried on by Tóth et al. [Tóth et al., 2017] suggested that individual differences in network topology are related to cortical maturation during the prenatal period. Tokariev et al. [Tokariev et al., 2018] found that phase synchrony in frontally connected networks was correlated with newborn's neurological performance. Kuchenbuch et al. [Kuchenbuch et al., 2019] proposed EEG multivariate indexes as biomarkers for the detection of specific aetiologies behind neonatal seizures.

This short background introduces the aims of the present work: evaluating different multivariate EEG parameters as indices for the characterization of neonatal seizures during the various stages of the seizure event. Moreover, this work aims at evaluating if the same indexes could allow discriminating between newborns with seizures and seizure-free ones. An evaluation of the trends of the seizure dynamics with respect to seizure-free EEG background is proposed, combining synchronicity phase indices and measure of stability of the global synchronized system (i.e. all the EEG channels). As suggest by Jiruska et al. [Jiruska et al., 2013], seizure are considered as a dynamic disorder whose properties constantly evolve over time.

3.2 Methods

All the proposed methods were implemented on the Helsinki Dataset, already introduced in Section 2.3 of this PhD thesis. Only the newborns who received unanimous consensus among the three experts were considered. Thus, the 33 newborns with consensus seizure events and the 22 seizure-free ones with unanimous consensus were taken into account. The AND combination was used to merge the experts' labels. For each EEG recording the following EEG montage was considered: Fp2-F4, F4-C4, C4-P4, P4-O2, Fp1-F3, F3-C3, C3-P3, P3-O1, Fp2-F8, F8-T4, T4-T6, T6-O2, Fp1-F7, F7-T3, T3-T5, T5-O1, Fz-Cz, Cz-Pz. Therefore, multichannel EEG signals are considered of size $n \times N$, where n=18 are the derivations and N are the samples in each derivation. The sampling frequency was 256 Hz. In order to perform network analysis, each one of the 18 EEG derivations was considered as a "node" of the brain network and any pair of EEG derivations (i.e. the "edges" of the networks) as a functional link. In network theory, nodes and edges are the essential building blocks of networks [Fornito et al., 2016]. To evaluate the properties of the brain networks the following parameters were considered:

- Stability of the global synchronized system [Schindler et al., 2008]. The Synchronizability (S) is described in the subsection 3.2.1.
- Measure of the degree of phase synchrony using the Circular Omega Complexity index (COC) [Baboukani et al., 2019] described in the subsection 3.2.2.

Both S and COC were calculated for the whole length of the EEG recordings on 2 seconds epochs with overlap of 1 second (50%) [Schindler et al., 2008]. A total of about 39'000 seizure windows and 363'000 seizure-free windows was analysed. Thus, EEG sub-windows of size 18 × 512 were created. A band-pass FIR filter between 0.5 Hz and 30 Hz was applied to each sub-window. Windows were then normalized (zero mean and unit variance). The proposed methods were implemented under MATLAB computing environment version 2019b (MATLAB and Statistics and Machine Learning Toolbox Release, 2019b).

3.2.1 Synchronizability index S

Although the synchronization of multivariate systems has been deeply investigated in neuroscience, due to the different approaches proposed a unique and rigorous definition of synchronization does not exist. In this study, Synchronization was defined as: "*a process whereby two (or many) dynamical subsystems adjust some of their time-varying properties to a common behaviour due to coupling or common external forcing*" Carmeli et al. [Carmeli et al., 2005]. Moreover, in network analysis, Synchronizability (S index) describes the degree of stability of the system's globally synchronized state [Comellas and Gago, 2007, Lehnertz et al., 2017]. This information could be used to characterize the time-varying brain dynamics using EEG recordings [Lehnertz et al., 2017]. In order to compute the S index, the approach proposed by Schindler et al. [Schindler et al., 2008] was applied here. First the cross-correlation R_{xy} (Equation 3.1) was applied to quantify the degree of connectivity between EEG derivations [Quiroga et al., 2002]:

$$R_{xy} = \begin{cases} \sum_{n=0}^{N-m-1} x_{n+m} y_n^* & m \geq 0 \\ R_{xy}^*(-m) & m < 0 \end{cases} \tag{3.1}$$

R_{xy} provides a measure of similarity between two signals x and y shifted by a delay factor m. For each couple of bipolar derivations, the value ρ_{xy}^{max} was calculated, defined as the maximum of the normalized cross-correlation in absolute value (Equation 3.2).

$$\rho_{xy}^{max} = max_n \left\{ \frac{|R_{xy}(m)|}{\sqrt{R_{xy}(0) R_{yx}(0)}} \right\} \tag{3.2}$$

Thus using 18 derivations a connectivity matrix C_{ij} of size 18x18 is obtained, whose values vary between 0 and 1. Since the main diagonal contains the autocorrelation values, these self-connections were excluded, setting the values to 0. C_{ij} provides a synthetic description of the functional connectivity between all the EEG signals considered. Each element of C_{ij} provides information about the similarity between each couple of derivations. Thus, the analysis of the properties of the whole matrix gives information about the synchronization among all the oscillators considered (i.e. the EEG derivations). The evaluation of possible effects of volume conduction on the cross-correlation values is provided in [Frassineti et al., 2021a, Supplementary Materials]. Then, from C_{ij} the adjacency matrix A_{ij} was computed to detect the main connectivity relationships between the derivations [Fornito et al.,

2016]. To remove spurious links and enhance the relevant topological properties of the network, a threshold must be applied to C_{ij}. The selection of the proper threshold is a critical point in network analysis [Fornito et al., 2016]. To overcome this limit, before binarization we applied to C_{ij} the disparity filter proposed by Serrano et al. [Serrano et al., 2009]. The filter gives a higher weight to the connections with higher statistical relevance. A significance level α is defined, that represents the degree of statistical confidence. The filter tests the following null hypothesis, for each element of C_{ij}: "*the normalized weights* (ρ_{xy}^{max}) *that correspond to the connections of a certain node of degree k are produced by a random assignment from a uniform distribution*" [Serrano et al., 2009]. The degree *k* of a node is defined as the number of adjacencies connected to the node.

The disparity filter finds all the relevant edges whose weights satisfy the relation:

$$\alpha_{ij} = 1 - (k-1)\int_0^{p_{ij}} (1-x)^{k-2} dx < \alpha \tag{3.3}$$

Where k is the degree of the node, $p_{ij} = \frac{w_{ij}}{s_i}$, w_{ij} is the weight of the i-th connection to its j-th adjacency, and s_i is the strength of the node i-th [Fornito et al., 2016]. α_{ij} is the probability that an edge of a given node satisfies the null hypothesis. To select the optimal value of α for each C_{ij} the following metrics were considered:

- the normalized clustering coefficient C/C_r [Fornito et al., 2016]. Where C is the clustering coefficient, that is a measure of local connectedness for C_{ij}. C_r is the clustering coefficient for a corresponding random network "*with a preserved degree distribution and an identical average number of edges per node*" [Schindler et al., 2008]. The degree distribution is the probability that a randomly selected node in the network has exactly *k* adjacency.
- the normalized average shortest path length L/L_r [Schindler et al., 2008]. Where L is the average shortest path length that provides a global measure of brain network's capacity to integrate information using the shortest path routing [Ponten et al., 2007]. As for C_r, L_r is the average shortest path length for a corresponding random network.
- the number of networks without any significant connection after the application of the disparity filter (hereinafter NetEmpty), i.e. an A_{ij} with all the elements equal to 0.

Values of C/C_r and L/L_r close to 1 correspond to networks with a random behaviour. High values of NetEmpty represent heavy pruning performed by the disparity filter on the networks. Table 3.1 shows the tests performed to obtain the parameter α, based on the considered metrics.

Table 3.1. Values obtained for C/C_r, L/L_r and NetEmpty with varying α for the 39 patients with consensus seizures. [Frassineti et al., 2021a].

C/C_r		***Seizure***		***Non-Seizure***	
α		Mean	Std	Mean	Std
0.4	C/C_r	1.24	0.15	1.20	0.13
0.35	C/C_r	1.93	1.28	1.81	1.19
0.3	C/C_r	3.02	2.70	2.98	2.67
L/L_r		***Seizure***		***Non-Seizure***	
α		Mean	Std	Mean	Std
0.4	L/L_r	1.01	0.03	1.01	0.03
0.35	L/L_r	1.08	0.14	1.06	0.13
0.3	L/L_r	1.02	0.30	1.01	0.28
NetEmpty		***Seizure***		***Non-Seizure***	
α		Absolute	Relative	Absolute	Relative
0.4	NetEmpty	0	0	5759	1.59%
0.35	NetEmpty	31	<0.1 %	6076	1.67%
0.3	NetEmpty	1488	4%	9455	2.60%

The average values of C/C_r, L/L_r and NetEmpty and their standard deviation obtained for the 39 patients with seizures are reported, during the seizure events and the non-seizure periods. According to [Serrano et al., 2009] an optimal range for α can be identified as 0.3< α < 0.4. For our methods we consider α = 0.35 as the best value to binarize C_{ij} obtaining A_{ij}. Indeed, this value preserves almost all the seizures in the dataset for the subsequent analysis.

In Figure 3.1 a non-seizure example is reported. The Cij matrix is shown in Figure 3.1b, where each element represents the ρ_{xy}^{max} value for each couple of bipolar derivations. The corresponding 2 s window of the normalized EEG signal is shown in Figure 3.1a. From C_{ij} the adjacency matrix A_{ij} was computed to detect the main connectivity relationships between the derivations.

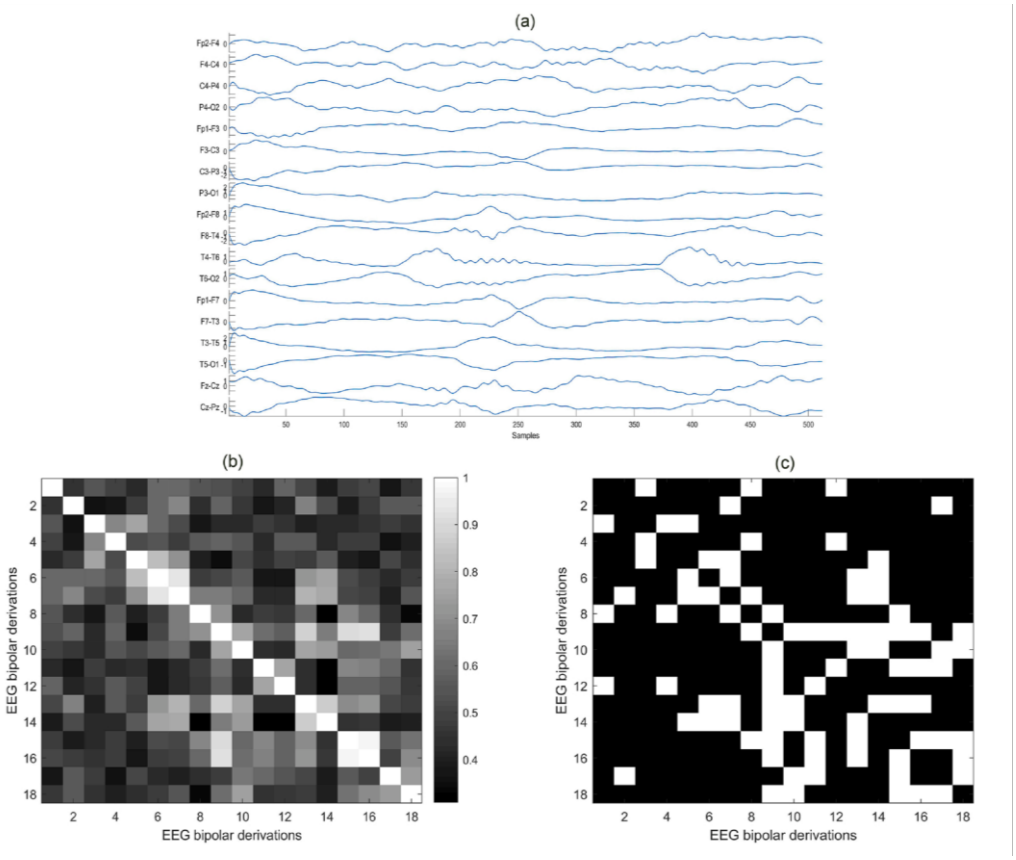


Figure 3.1. (b) Example of Connectivity Matrix C_{ij} derived from the normalized EEG window (a). The matrix A_{ij} after the binarization process is shown in (c). [Frassineti et al., 2021a].

Finally, from A_{ij} its Laplacian matrix $\Lambda_{ij} = k_i\delta_{ij} - A_{ij}$ was computed, where δ_{ij} is the Kronecker delta and k_i is the degree of the node i-th [Fornito et al., 2016]. Thus, as in [Schindler et al., 2008], the index of Synchronizability (S) can be defined as in Equation 3.4:

$$S = \frac{\lambda_{max}}{\lambda_{min}} \tag{3.4}$$

Where λ_{min} and λ_{max} are the smallest non-zero eigenvalue and the maximum eigenvalue of Λ_{ij} respectively. S describes the stability of the globally synchronized state. It provides a quantitative measure of the degree of stability for different systems: the system with high values of S tends to instability, thus losing its properties of synchronization, while the system with low values S tends to be synchronizable [Schindler et al., 2008]. Therefore, S may be useful to compare the trend of a seizure with that of other events or different seizures. For these reasons, S was considered in this

work as the first parameter to discriminate between neonatal seizure events and non-seizure periods or seizure-free patients. The quantitative and qualitative results of this analysis will be presented in Section 3.3.

3.2.2 Circular Omega Complexity (COC) as measure of phase synchrony

Recently, Baboukani et al. [Baboukani et al., 2019] proposed Circular Omega Complexity (COC) to measure the phase synchrony in EEG recordings of newborns. They used COC to detect neonatal seizures as well as to distinguish between burst and interburst periods. Given such promising results in this work it has been evaluated if COC might be used to characterize the seizure's network dynamics in the newborn. According to [Baboukani et al., 2019], the Stationary Wavelet Transform (SWT) with 6 level of decompositions was applied to each 2 seconds EEG epochs and to all the D=18 derivations. The mother wavelet function was the Symlet2 (sym2) [Baboukani et al., 2019]. The size of each EEG derivation was 1xn, where n is the number of samples of each epoch. Thus, the SWT decomposition gave a matrix 7xn. As in [Frassineti et al., 2020], the detail levels from d1 to d3 were removed, being relative to high frequencies not useful for this analysis [Rankine et al., 2006], obtaining a matrix 4xn. The SWT decomposition allows separating the frequency components, from the highest frequencies for d4 (approximately between 8 and 16 Hz), to the lowest frequency in the approximation level a6 (0-2 Hz). To compute the COC index, for each EEG derivation and for each decomposition level of SWT, let $x(n)$ the detail or approximation level obtained by SWT. The Hilbert Transform was applied to $x(n)$, representing it into its analytical form $z(n)$ as described in Equation 3.5:

$$z(n) = x(n) + i\hat{x}(n) = A(n)e^{i\varphi(n)} \tag{3.5}$$

Where $\hat{x}(n)$ is the Hilbert Transform of $x(n)$, $A(n)$ is the instantaneous amplitude and $\varphi(n)$ the instantaneous phase as shown in Equation 3.6:

$$\varphi(n) = \arctan\left(\frac{\hat{x}(n)}{x(n)}\right) \tag{3.6}$$

For all the D derivations and SWT levels, a matrix $\mathbf{X}$ ($D \times n$) is built: $\mathbf{X}=\{x_1(n), \dots, x_D(n)\}$, along with the corresponding $\varphi(n)$. Then the circular correlation $\boldsymbol{c}_{k,l}^{\boldsymbol{X}}$ is computed between two instantaneous phases $\varphi_k(n)$ and $\varphi_l(n)$, where k and l represent any pair of signals with fixed level of decomposition or approximation. $\boldsymbol{c}_{k,l}^{\boldsymbol{X}}$ can be defined as in Equation 3.7:

$$\boldsymbol{c}_{k,l}^{\boldsymbol{X}} := \frac{\sum_{n=0}^{N-1} \sin(\varphi_k(n) - \overline{\varphi_k}) \sin(\varphi_l(n) - \overline{\varphi_l})}{\sqrt{\sum_{n=0}^{N-1} \sin^2(\varphi_k(n) - \overline{\varphi_k}) \sin^2(\varphi_l(n) - \overline{\varphi_l})}} \tag{3.7}$$

Where $\overline{\varphi_k}$ is the circular mean of $\varphi_k(n)$:

$$\overline{\varphi_k} = \arg\left(\sum_{n=0}^{N-1} e^{i\varphi_k(n)}\right) \tag{3.8}$$

Then the circular correlation matrix CCM for $\mathbf{X}$ is defined as $CCM^{\boldsymbol{X}} := [\boldsymbol{c}_{k,l}^{\boldsymbol{X}}]_{D\times D}$, and λ_m are its eigenvalues with m=1…D. $CCM^{\boldsymbol{X}}$ provides information about the synchronization level among the elements in $\mathbf{X}$. Finally, it is possible to derive the COC index from $CCM^{\boldsymbol{X}}$ as described in Equation 3.9:

$$COC := 1 + \frac{\sum_{m=1}^{K} \overline{\lambda_m} \log \overline{\lambda_m}}{\log D} \tag{3.9}$$

Where $\overline{\lambda_m} = \frac{\lambda_m}{\sum_{m=1}^{D} \lambda_m}$. COC is a parameter with values between 0 and 1, where 0 corresponds to no phase synchrony between the considered elements in **X**. To get a single index describing the whole system, the mean value of all the COC values obtained for all the SWT levels was calculated. Therefore, the COC parameter obtained by Hilbert Transform and SWT was the second parameter used in this work for the characterization of neonatal seizures. The results obtained are presented in Section 3.3. To better understand how S and COC could characterize seizure's network dynamics, Figure 2a shows an example of a single seizure of 20 s of duration. Fig. 2b shows the trend of S during the seizure. High values of S mean high instability. They are at the onset of the seizure (between windows 1 and 3) and at the end (between windows 15–16). In its central part the seizure presents low values of S that means higher stability of the globally synchronized state as compared to the seizure's onset. Fig. 2c shows an example of the COC trend for the seizure shown in the upper plot. Notice the increasing value of COC in the initial phase, followed by a sort of plateau and decreasing values towards the seizure's end.

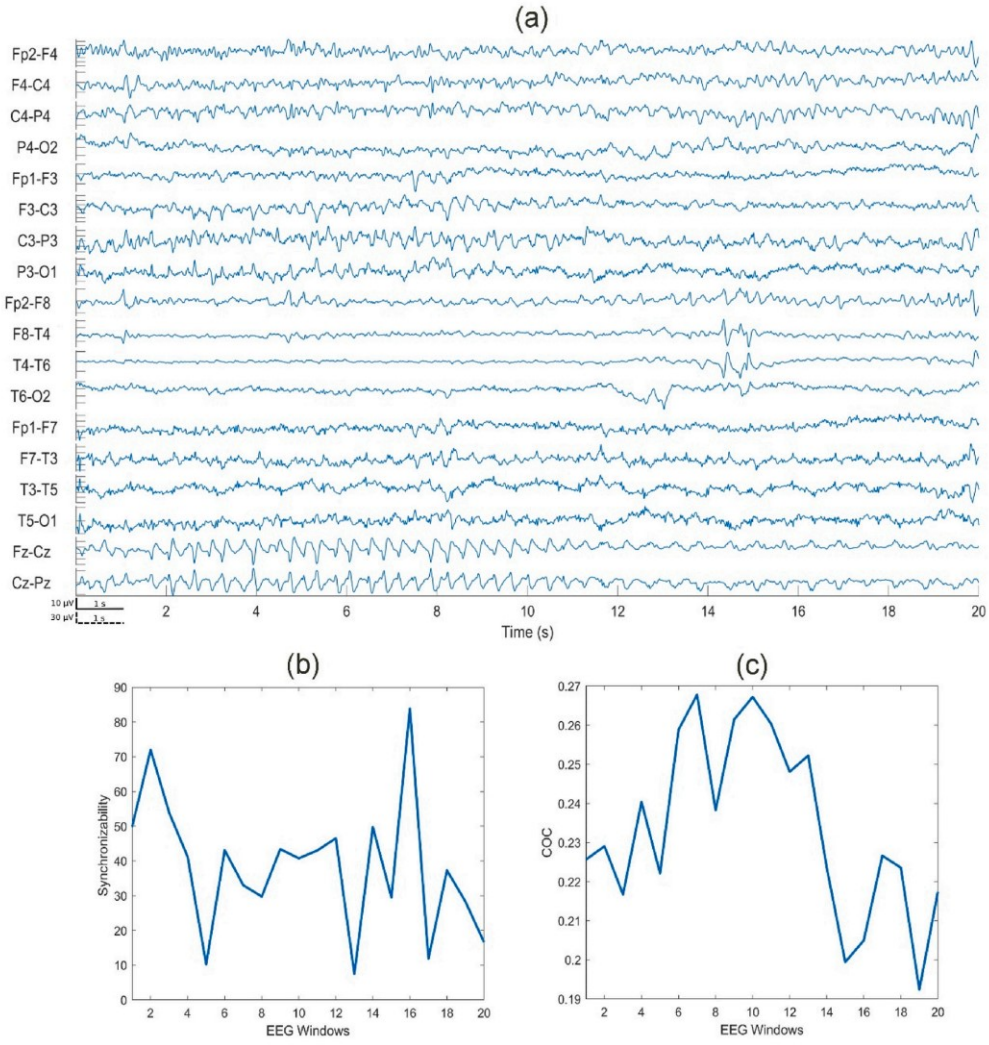


Figure 3.2. (a) EEG of a seizure lasting 20 s (y-axis μV). For the y-axis two different scales are used, shown at the bottom of (a): Solid bar: from Fp2-F4 to T5-O1; Dashed bar: for Fz-Cz and Cz-Pz. (b): Synchronizability values during the seizure. (c): COC trend during the seizure. [Frassineti et al., 2021a].

3.2.3 Statistical evaluation for S and COC trends

In this work the dynamics of seizure events was provided by COC and S measures, however it could be challenging to compare events with different time duration, as well as the number of seizure events that could vary among subjects. Therefore, the approach proposed by Schindler et al. [Schindler et al., 2008] was followed: both S and COC values of each seizure event were subdivided and re-sampled into a predefined number of bins, 10 in our case, and then compared to their pre- and post-seizure time instants. In this work pre- and post- seizure instants were defined as the 2s epochs immediately before and after the ictal event [Schindler et al., 2008]. A moving average filter was applied to obtain 12 fixed instants for each seizure events: 1 for the pre-seizure, 10 for the seizure, and 1 for the post-seizure. Each seizure trend was normalized between 0 and 1 to compare events with different amplitudes, preserving only their dynamics and allowing a uniform evaluation of all the seizure events of the Helsinki Dataset. Then a statistical analysis was performed to assess if the normalized and standardized seizure trends were different from the trends of seizure-free patients. On analogy to surrogate analysis performed in [Lancaster et al., 2018], three different tests were defined. To this aim, M random EEG segments were extracted from each of the 22 seizure-free patients. Then the following tests were performed on the EEG recordings:

- Test T1: M EEG random segments of length V=24s each, with random and non-consecutive samples.
- Test T2: M EEG random segments of length V=24 each, with random starting points and consecutive samples. As for T1, V=24s was selected in order to compare the results with the standardized bins defined for the patients with seizures, thus mimicking the 12 epochs of 2 seconds as in the patients with seizures.
- Test T3: M EEG random segments with different duration V=[V_{min} … V_{max}], random starting points and consecutive samples. Where V_{min}=30s and V_{max}=300s. They were chosen according to seizure duration in the Helsinki Dataset.

Several values for M were considered: M=10,50,100 and 500, obtaining a first plateau for the metrics with M=100, mitigating effects due to random fluctuations of amplitude values for S and COC. Thus, in section 3.3 only the results obtained with M=100 will be reported.

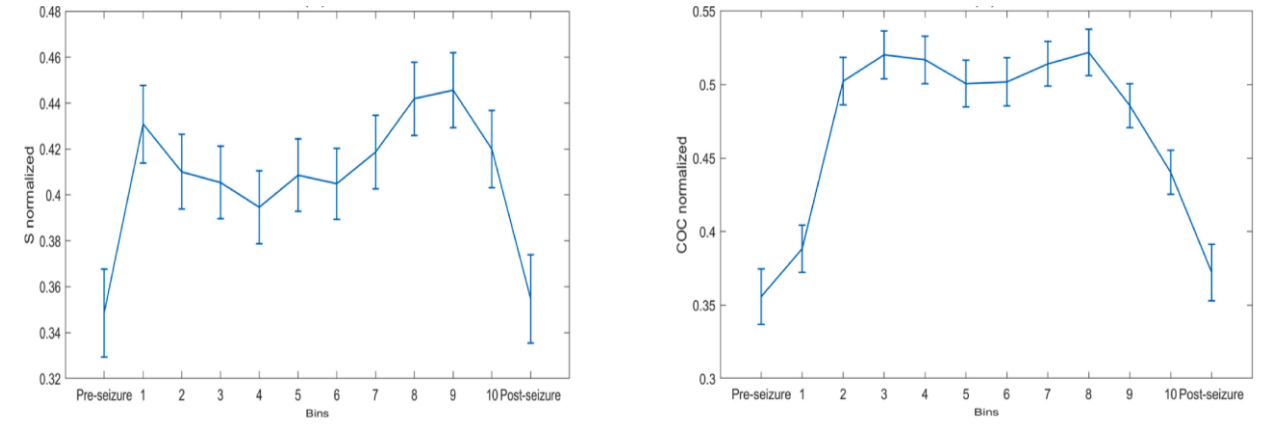
3.3 Results

In Table 3.2 the results of S and COC obtained from the 2s epochs of the 39 patients with consensus seizures are shown. Mean and standard values ($\mu \pm \sigma$) of S and COC for both seizure epochs and non-seizure epochs are reported. A Mann-Whitney test was applied in order to evaluate differences between seizure epochs and non-seizure epochs (significance level 0.05). Both the p-value and the Cohen's d are reported. As shown in Table 3.2 the COC parameter allowed a good discrimination between seizure epochs and non-seizure epochs (p-value<0.001 and high Cohen's d=0.86), instead S did not provide a good differentiation among epochs (p-value <0.001 but low Cohen's d= 0.1).

Table 3.2. Statistical Results of S and COC for the 39 patients with unanimous consensus for seizure events. Two-sided Mann-Whitney test was applied, considering all the 39000 seizure epochs and the 360000 non-seizure epochs.

	Seizure epochs	**Non-Seizure Epochs**	**Test Mann-Whitney**	
Metrics	$\mu \pm \sigma$	$\mu \pm \sigma$	p-value	Cohen's d
S (α=0.35)	19.43 ± 15.97	17.82 ± 14.55	p<0.001	0.1
COC	0.29 ± 0.07	0.25 ± 0.04	p<0.001	0.86

Furthermore, a supplementary statistical analysis was performed, considering each patient as a single test and comparing its seizure epochs with non-seizure epochs. Using the Benjamini-Hochberg correction for multiple comparisons, significant differences were found for S metrics in 24 patients (61%) and for COC metrics in 32 patients (82%). These findings confirmed what is shown in Table 3.2: COC metric is more reliable than S metric for the seizure detection and characterization problem. Figure 3.3 shows the normalized and standardized trends obtained for the metrics S and COC, generating them as proposed in [Schindler et al., 2008]. The plots show the average values and their standard error (normalized for all the 343 seizure events) obtained for the seizure's bins and the corresponding pre-seizure and post-seizure instants.



(a) (b)

Figure 3.3. Normalized and standardized trends of S (a) and COC (b) applied to all seizures. [Frassineti et al., 2021a]

Table 3.3 shows the results obtained for S and COC on the 22 seizure-free patients. As in Table 3.2, the average values and their standard deviation ($\mu \pm \sigma$) for both the metrics are reported for all the three tests described in section 3.2.3 (Test T1, T2 and T3 with M=100).

Table 3.3. Average and standard deviation values of S and COC for the 22 seizure-free patients for the three tests T1, T2 and T3 defined in section 3.2.3 using M = 100. [Frassineti et al., 2021a].

	T1	**T2**	**T3**
Metrics	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$
S (α=0.35)	12.65 ± 8.47	14.72 ± 9.90	14.79 ± 10.17
COC	0.19 ± 0.02	0.25 ± 0.04	0.26 ± 0.04

Table 3.3 points out that seizure-free patients show values close to the non-seizure epochs of the patient with seizure events, especially for the COC metric, where COC average values are equal to 0.25 for non-seizure windows and to 0.26 for test T3 with M=100, respectively.

Finally Figure 3.4 shows the trends obtained for the seizure-free subjects for the tests T1, T2 and T3 (M=100, solid lines). For a direct comparison the S and COC trends obtained for the seizure events are shown in the same figures (dashed lines). To facilitate the comparison, the first and the last bin of the seizure-free trends are called pre- and post-seizure bins. Obviously, they do not correspond to real pre- or post-seizure events, but just to the windows immediately before and after the seizure-free trends. Odd drops in Figure 3.4 for the seizure-free case may be due to the standardization and normalization step. Indeed, in this case such pre-processing might enhance very small differences, resulting in such drops. A comprehensive evaluation of this behaviour is reported in [Frassineti et al. 2021a].

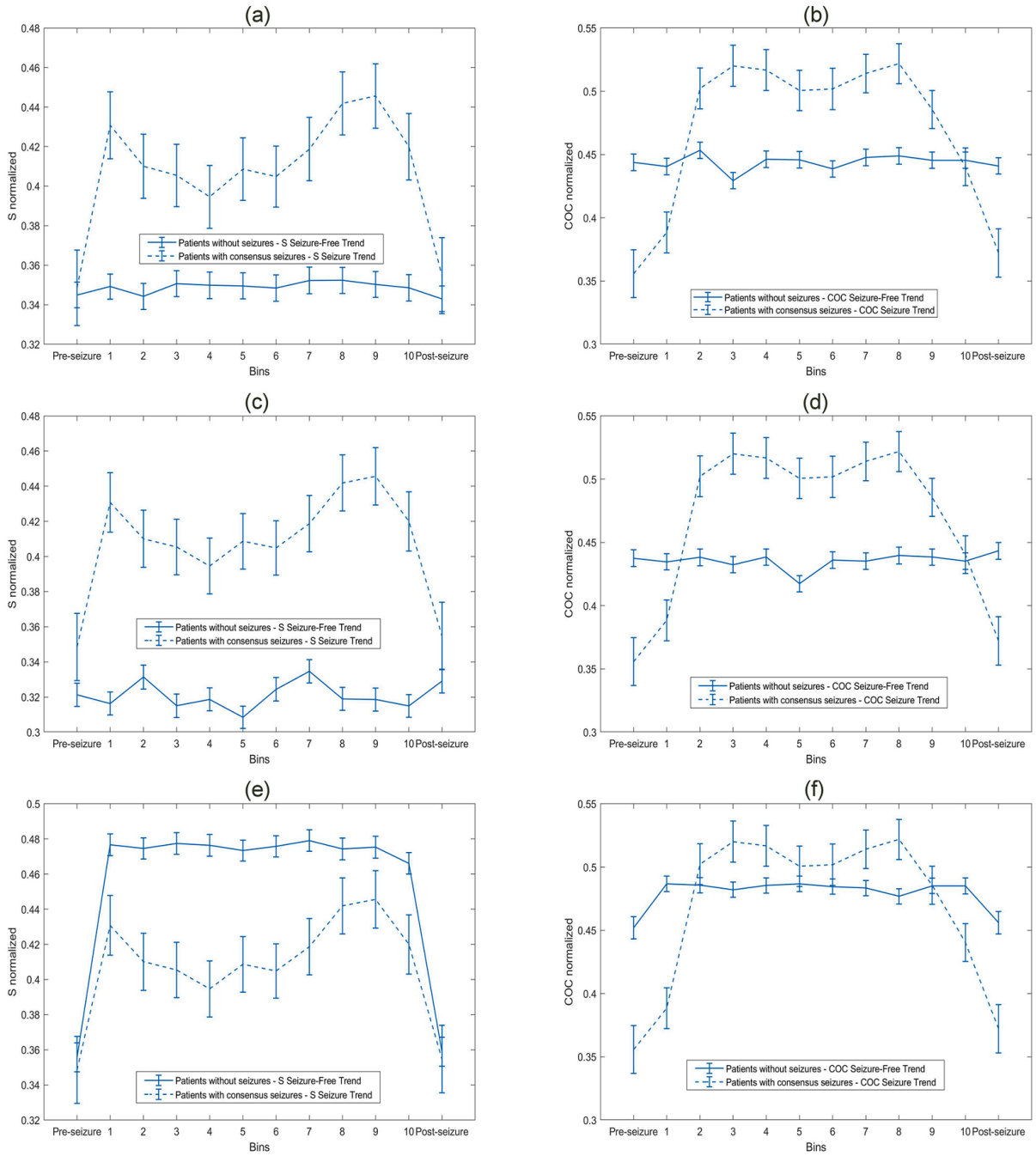


Figure 3.4 (a,b) Normalized and standardized trends of S and COC for test T1; (c,d) the same for test T2; (e,f) the same for test T3 (M = 100). The seizure-free trends (solid lines) are compared to the normalized seizure trends (dashed lines) shown in Figure 3.3. [Frassineti et al., 2021a].

However, the representation by standardized trends, as in Figures 3.3 and 3.4, might hide specific details for some seizures, especially for those of long duration. In future developments, a multiple level representation could be evaluated.

Another important point is evaluating the transition from the pre-ictal periods to ictal periods and from ictal to post-ictal periods. Thus, in this work, it has been evaluated if S and COC metrics could be able to detect transitions from pre and post-ictal periods. Figure 3.5a shows the average evolution of COC metrics starting from 10 epochs before the first 5 epochs of all the seizure events (index from -10 to -1). The "0" window is the first seizure epoch, i.e. its onset, and is denoted by arrows. Figure 3.5b shows the last five epochs of all the seizures (index from -4 to 0). The "0" window represents

the offset of the seizures and is indicated by arrows. The same illustrative analysis was proposed for S metrics, and it is shown on Figure 3.5c and 3.5d. Stars denote significant changes ($p < 0.05$; one-sided Wilcoxon signed-rank test).

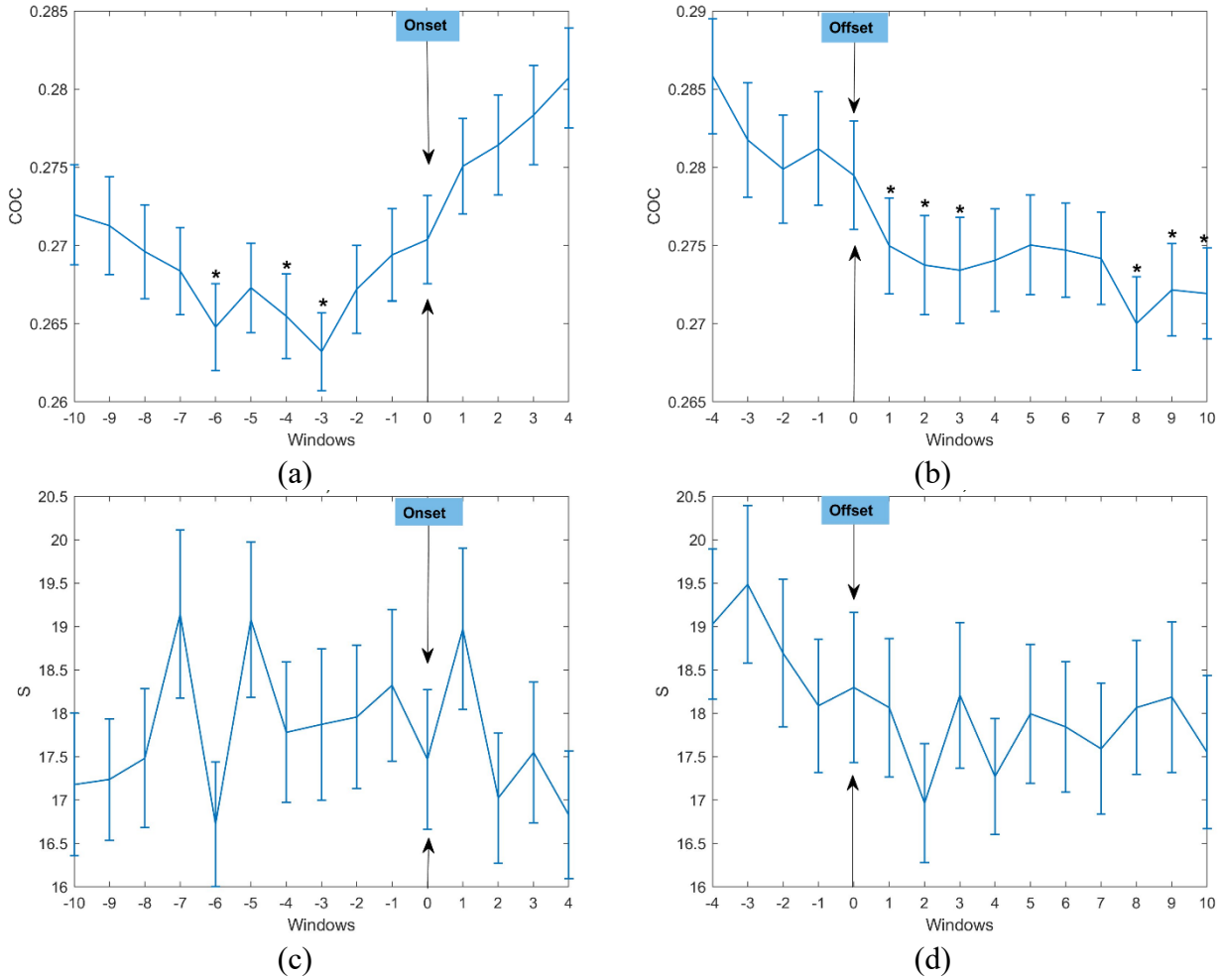


Figure 3.5. (a) The COC trend from the 10th window before the onset of the seizure (windows 0) to the 4th window after the onset. Stars denote significant changes between fist ictal window 0 and the specific pre-ictal window ($p < 0.05$; one-sided Wilcoxon signed-rank test). (b) The COC trend from the 4th window before the offset of the seizure (windows 0) to the 10th window after the end of the seizure. Stars denote significant changes between last ictal window 0 and the specific post-ictal window ($p < 0.05$; one-sided Wilcoxon signed-rank test). (c) The S trend from the 10th window before the onset of the seizure (window 0) to the 4th window after the onset. (d) The S trend from the 4th window before the offset of the seizure (window 0) to the 10th window after the end of the seizure. The figure shows the average values obtained from all the seizures (330 seizures lasting at least 10 s) and their standard errors obtained after normalization of all the seizure events considered. [Frassineti et al., 2021a].

To evaluate differences between a single window and the whole trend, for both metrics a one-sided Wilcoxon signed rank-test (significance level 0.05) was performed between the fist ictal window and each pre-ictal window. The same analysis was performed on post-ictal transitions. Results of this analysis are reported in Table 3.4. For COC values, during the pre-ictal transition there are significant differences between the first ictal window and pre-ictal windows around windows -6 and -3. Instead for the post-ictal transition, with COC metrics, differences on windows 1 and 3 and from windows 8-10 are found. Instead, S metrics did not show any significant differences both for pre- and post-ictal transitions.

Table 3.4. Results of one-sided Wilcoxon test (signifiance level 0.05) between the fist seizure windows and pre-ictal windows (Pre-Ictal Transition) and between the last seizure windows and the post-ictal windows (Post-Ictal Transition). Both COC and S metrics are reported. A total of 330 consensus seizures were considered. [Frassineti et al., 2021a]

COC				S ($\alpha = 0.35$)			
Pre-Ictal Transition		Post-Ictal Transition		Pre-Ictal Transition		Post-Ictal Transition	
Window	p-value	Window	p-value	Window	p-value	Window	p-value
−10	0.76	1	0.02	−10	0.34	1	0.67
−9	0.62	2	0.01	−9	0.56	2	0.35
−8	0.23	3	p<0.01	−8	0.67	3	0.59
−7	0.32	4	0.05	−7	0.91	4	0.52
−6	p<0.01	5	0.10	−6	0.44	5	0.44
−5	0.07	6	0.13	−5	0.91	6	0.46
−4	0.02	7	0.05	−4	0.68	7	0.41
−3	p<0.01	8	p<0.001	−3	0.53	8	0.71
−2	0.16	9	0.02	−2	0.85	9	0.40
−1	0.28	10	0.03	−1	0.86	10	0.16

3.4 Discussion and Conclusions

The work presented here evaluated if two different properties of brain's network, i.e. the stability and the phase synchronicity of brain's network dynamics, derived from EEG recordings, could be helpful to characterize neonatal seizures. The results in Table 3.3 and Figure 3.3 showed that the proposed index of stability, i.e. the S metrics, exhibited a rapid growth at the initial bins of the seizure trends, corresponding to higher instability of the brain network. At the same time the networks showed increasing values of COC, remaining almost constant during the central part of the seizure trend (Figure 3.3b, from bins 2 to 8). S trend seems to be more stable during the central part of the seizures (Figure 3.3a, bins from 3 to 6). Instead, at the end of the seizures the networks seem more unstable (high values of S), at the same time losing their high phase synchronization (decreasing values of COC, Figure 3.3b, bins from 7 to 9).

These results suggested an interpretation of what might happen on the brain's network during a neonatal seizure: close to the seizure's onset there is a first phase of high instability, during the central part the networks exhibit a high synchronicity and stability, then close to the offset the brain's networks shows an instability configuration, losing their synchronization properties. These physiological considerations are partially in accordance with the literature related to the seizure's dynamics in the adult [Jiruska et al., 2013; Breakspear et al., 2006; Gray and Robinson, 2009].

As proposed in [Gray and Robinson, 2009], stability reflects the structure of the brain's networks in terms of number of connections and their average connection gains. Any small variation in gains can decrease the network's stability and such variations could be found during the transition from and to an ictal phase [Jiruska et al., 2013]. As reported in [Baboukani et al., 2019] and as shown in Table 3.3, the high values of COC might reflect the very high firing of neurons during a seizure and thus the higher degree of connections between different areas of the brain network. This activity tends to saturate the capacity of neurons to fire and some brain areas may become silent close to the offset [Schindler et al., 2008]. Thus, it implies a fragmentation of the brain network and a decrease of global synchronizability (i.e. the S metrics) and consequently a progressive loss of synchronization. However, as discussed by Breakspear et al. [Breakspear et al., 2006], these considerations must take into account that during a seizure other physiological parameters may vary, thus other mechanisms could control the evolution or the end of the seizure, especially in the case of neonatal seizures.

Results in Table 3.3 and 3.4 confirmed that COC metrics could be used as multichannel features of NSD, as already suggested by Baboukani et al. [Baboukani et al., 2019]. Instead, S metrics did not give reliable enough results to propose it as possible multichannel feature for NSD. The results of

Test 1, 2 and 3 reported in Table 3.3, confirmed the differences between patients with seizure and seizure-free ones. This means that the seizure's trend is related to specific brain's network dynamics and not due to random behaviours.

As shown in Table 3.4, the analysis of COC metrics values might suggest the presence of some significant changes in the trend related to a possible anticipation of the seizure event. However, as only the consensus annotations were taken into account, this could reduce the duration of the seizures. On the other hand, defining the true seizure onset is a difficult task, especially in the case of newborns. Thus, in order to confirm these findings further investigations are needed. For example, the approach proposed in [Schindler et al., 2006] could be investigated for the definition of the seizure onset/offset, while another approach could consider different combinations of the experts' annotations.

In conclusion, results show that during the seizures the metrics reflect the evolution of the network dynamics: the seizure's network dynamics significantly alter the metrics more than the non-seizure periods. However, the proposed metrics cannot be considered as a direct or exclusive representation of the seizure dynamics, because artefacts coming from extra-cerebral sources, or non-seizure sources may occur at the same time. Indeed, a neonatal seizure can often be a focal one. Therefore, further analyses are required to discriminate between the seizure dynamics, the extra-cerebral sources and the non-seizure sources, to extract the specific seizure dynamics that might occur during the ictal event. Moreover, the results were tested on one public dataset only. To confirm the findings and the repeatability of the proposed procedure further testing on different dataset must be performed.

Another point of investigation would be the use of the source space instead of the sensor space for the functional connectivity analysis. However, as stated in [Christodoulakis et al., 2013], accessing the source reconstruction and its quality depends on the number of available electrodes that usually are quite few in NICU. Therefore, the sensor space might be the only one that can be used in practice. Moreover, in this work 18 EEG derivations were considered, thus further investigation regarding the minimum number of derivations (and which one) required for reliable multichannel EEG features as S and COC could be performed in future developments.

Given the promising results, the methods might be extended and tested on other neurological applications where fast and reliable automatic analysis of EEG is required. For example, they might be helpful to characterize burst and interburst over time in preterm newborns [O'Toole et al., 2017], or to evaluate the network dynamics of pre-seizure trends [Liu et al., 2018]. The proposed methods could also support already existing seizure detectors as a post-processing procedure on alleged seizures detected by the systems, the interpretation of standardized trends of S and COC being useful for further evaluation and validation.

4. The role of Autonomic Nervous System and Heart Rate Variability in neonatal seizure detection and characterization.

Some contents in this chapter are based on the following publications:

- *Frassineti, L., Lanatà, A., Olmi, B., Manfredi, C., 2021. Multiscale Entropy Analysis of Heart Rate Variability in Neonatal Patients with and without Seizures. Bioengineering, Vol. 8, Issue 9, p. 122. https://doi.org/10.3390/bioengineering8090122*
- *Frassineti, L., Manfredi, C., Ermini, D., Fabbri, R., Olmi, B., Lanatà, A., 2022. Analysis of Brain-Heart Interactions in newborns with and without seizures using the Convergent Cross Mapping approach. In 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp. 36-39, doi: 10.1109/EMBC48229.2022.9871141.*

As already stated in Chapter 2, the involvement of the Autonomic Nervous System (ANS) during or close to neonatal seizures is well known in literature. However, the complex dynamics and interactions between ANS and the Central Nervous System (CNS) during ictal events remain still unclear in the newborn. If such behaviours will be explained or characterized, they may be used to develop NSD even without the use of EEG techniques, allowing the monitoring of newborn with seizure events even in neonatal settings different from Neonatal Intensive Care Units (NICUs). As an example, the activity of ANS can be described by Heart Rate Variability analysis, obtainable by techniques simpler and more affordable than EEG, such as Electrocardiography (ECG) or Photoplethysmography (PPG). However, in literature there is still a lack of information about which HRV features could be helpful for the characterization and detection of neonatal seizures. To this aim, in this work a first evaluation about the use of multiscale HRV entropy features to discriminate between newborn with seizures and seizure-free ones is provided.

The chapter is organized as follows: in Section 1 a short introduction about the evidence in literature regarding the involvement of ANS during or close to neonatal seizures is given; Section 2 describes which entropy features were used in this work. Results of multiscale HRV entropy analysis are shown in Section 3, and the discussion about them is provided in Section 4. Moreover, considering the promising results obtained, in Section 5 a first analysis about the Brain-Heart Interactions (BHI) during neonatal seizures is provided, estimated by the Convergent Cross Mapping (CCM) approach. This further evaluation was performed to verify and quantify the possible interactions between CNS and ANS, evaluating differences of BHIs between newborns with seizure events and seizure-free ones. The findings obtained with multiscale HRV entropy features might be related to such different interactions during ictal events.

4.1. Introduction: Autonomic Nervous System, HRV and neonatal seizures

The detection of neonatal seizures is still tricky and time-consuming, even in highly specialized settings such as neonatal intensive care units (NICUs). Due to different manifestations, ictal patterns and aetiologies, their identification based only on the observation of the clinical signs is challenging [Pressler et al., 2021, Malone et al., 2009]. Currently, electroencephalography (EEG) is the accepted gold standard often combined with synchronized video recordings (video-EEG) [Pressler et al., 2021]. However, as already discussed in Chapter 2 and 3, the use of EEG or video-EEG requires expert staff available 24/24 h for proper detection and interpretation of ictal events, a condition often not

practicable for several neonatal settings [Dilena et al., 2021]. Indeed, a recent Italian survey showed that only 47% of neonatology centers have access to NICU services, and at least 20% of them cannot activate a 24h EEG monitoring [Mosca et al., 2019]. In a recent survey proposed by Wang et al. [Wang et al., 2021], 64% out of 251 hospitals provide EEG in NICUs. Moreover, about 10% of NICUs have no access to any EEG-based monitoring (EEG nor amplitude-EEG (aEEG)), and almost the 21% of NICUs have only aEEG [Boylan et al., 2010].

To overcome this issue, several EEG-based neonatal seizure detectors (NSDs), including artificial intelligence techniques, have been introduced to support the physician's decision [Olmi et al., 2021]. At the same time, there has been increasing interest on investigating other signals besides EEG for neonatal seizure detection and characterization [Olmi et al., 2021]. In particular, electrocardiography (ECG) might be a valid support to this task thanks to its ease of use, less invasiveness and lower costs than EEG. Therefore, ECG-based methods could be useful alternatives when EEG-related techniques are not readily achievable.

The latest research findings showed that newborns' autonomic nervous system (ANS) response is directly or indirectly involved during seizure events [Statello et al., 2021]. In particular, the Central Autonomic Nervous System (CANS) or Central Autonomic Network (CAN) may have a relevant role in seizure onset. Similarly to the "epileptic heart" in the adult and the child [Akyüz et al., 2021], CAN controls information from and to the heart involving also cortical, subcortical and brain stem regions. Thus, a disorganized electrical discharge, such as a seizure, might cause an autonomic dysfunction [Statello et al., 2018].

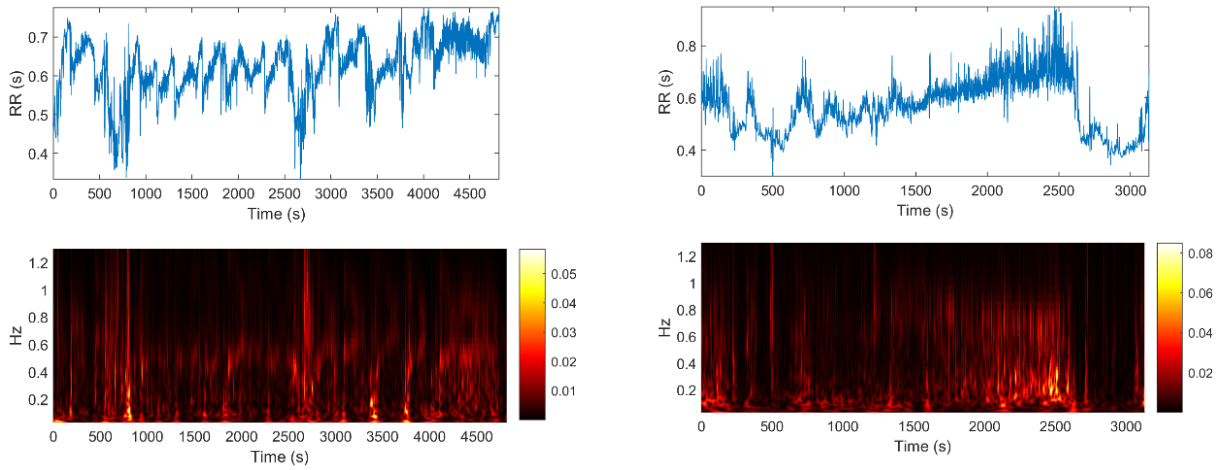
From ECG, the heart rate variability (HRV) analysis could provide helpful information about neonatal seizures, considering that HRV analysis could reflect the ANS activity [Shaffer and Ginsberg, 2017]. In particular, using HRV frequency measures, in [Statello et al., 2018] it was shown that seizures might impact the autonomic cardiovascular regulation in newborns. It has been proved that Hypoxic Ischemic Encephalopathy (HIE), the most common aetiology behind neonatal seizures [Pressler et al., 2021], is related to altered HRV signals [Bersani et al., 2020]. These results suggest that nonlinear HRV measures could help to untangle the complex mechanisms between neonatal seizures and the autonomic system. Information theory (IT) and nonlinear analysis methods might provide reliable information about neonatal seizure dynamics and characteristics.

In particular, entropy features used for HRV analysis in the adult were promising for detecting autonomic systems variations [Humeau-Heurtier, 2020]. Multiscale entropy analysis could increase the information obtained with single scale analysis about the underlying physiological process [Costa et al., 2005]. Thus, taking into account the differences between ictal events in adults and infants both for aetiology and the electro-clinical characteristics [Pressler et al., 2021], the application of these findings to newborns could be usefully exploited.

This study aims at evaluating whether HRV multiscale entropy features might provide helpful information for characterizing autonomic system dysregulation during neonatal seizures. Moreover, we investigated if this information could be helpful in neonatal seizure detection. In particular, the effectiveness of entropy features in the discrimination between newborns with seizure events and seizure-free ones have been explored, determining at which scale these differences become evident.

4.2. Multiscale HRV entropy methods

All the proposed methods were implemented on the Helsinki Dataset, already introduced in Section 2.3 of this PhD thesis. Only the newborns who received unanimous consensus among the three experts were considered. Thus, the 39 newborns with seizure events and the 22 seizure-free ones with unanimous consensus were taken into account. However, 9 patients (6 with seizures and 3 seizure-free) were excluded because the ECG signal was not present in their recordings or was highly corrupted by noise. Thus, the analysis has been performed on 33 patients with seizure events and 19 seizure-free patients. As for the EEG channels, the ECG signals were recorded using the NicoletOne vEEG System, Natus Medical [Frassineti et al., 2021d]. Two leads were placed on the newborn's chest, and the ECG was acquired with a sampling rate of 256 Hz. For each ECG recording, the inter-beat-interval (IBI) time series has been obtained using the Kubios software (Version 3, Kubios Oy, Kuopio, Finland) [Tarvainen et al., 2014], a widely used tool for the HRV analysis and the Pan-Tompkins' method for IBI time series extraction [Pan and Tompkins, 1985]. In figure 4.1 an example is shown of two RR trends (whole recordings). Figure 4.1a represents a trend from a seizure-free newborn, while Figure 4.1b a newborn with seizure events.



(a) **(b)**

Figure 4.1. (a) Newborn with seizure events. (b) Seizure-free subject. Above: inter-beat (RR) time series extracted by Kubios. Below: the corresponding time-frequency representation using Continuous Wavelet Transform. The frequency range is 0.04–1.3 Hz [Frassineti et al., 2021d, Statello et al., 2018].

Since Kubios software includes a limited number of nonlinear HRV measures, the IBI time series were exported and all the multiscale entropy HRV analysis and relative statistical analysis were performed under MATLAB 2020b environment.

The HRV entropy features were calculated on the segmented IBI time series applying a sub-windowing procedure, with non-overlapping windows lasting 4 min [Lucchini et al., 2016]. These procedures were repeated for each patient for the entire recording and discarding the windows that included recording pauses [Stevenson et al., 2019].

Furthermore, each window was labelled according to the information provided by the experts. Three classes have been selected: class "1" identifies a window with seizure events lasting at least 1s; class "int" (i.e., interictal) identifies a window without seizure events, but belonging to a patient with seizures; class "0" identifies a window without any seizure event and belonging to a seizure-free patient. The AND combination was used to merge the experts' labels. In other words, we defined as

"seizure window" only that for which the three experts simultaneously labelled that window as a seizure. A total of 441 interictal windows, 284 windows with seizure events, and 342 seizure-free windows have been achieved. Thus, a total of 1067 windows from the 52 patients were obtained.

For each window, the following HRV entropy measures were computed: Approximate Entropy (AE) [Pincus, 1991]; Sample Entropy (SE) [Richman and Moorman, 2000]; Generalized Sample Entropy (GSE) [Costa and Goldberger, 2015] and Fuzzy Entropy (FE) [Chen et al., 2007]; Permutation Entropy (PE) [Bandt and Pompe, 2002] and Distribution Entropy (DE) [Li et al., 2014]. All the indexes were evaluated at different scales, defining a coarse-grained time series for each of them.

4.2.1 Multiscale entropy analysis and the coarse-grained procedure

Research findings showed that multiscale scale entropy indexes in HRV analysis are helpful both for the characterization or detection of several pathologies and the description of different dynamics of the autonomic system [Costa and Goldberger, 2015]. Recently, the multiscale entropy analysis was successfully applied to ECG signals to detect and characterize seizures in adults [Humeau-Heurtier, 2020].

These results open the possibility of assessing the reliability of these techniques also in newborns. This work implements the multiscale analysis for all the entropy indexes (except for GSE) using the mean operator's coarse-grained procedure to generate the time series at each scale. The coarse-grained procedure is reported in Equation 4.1:

$$y^s(j) = \frac{1}{s} \sum_{i=(j-1)s}^{js} x(i) \quad , 1 \leq j \leq \lfloor N/s \rfloor \tag{4.1}$$

Where y is the coarse-grained time series, s is the scale factor, and $x(i)$ is the i-th sample of the original time series of length N samples. The symbol $\lfloor . \rfloor$ indicates the integer part of its argument.

Figure 4.2 shows the coarse-grained procedure from scale 2 to scale K. Note that, at scale 1 (s=1), the coarse-grained series is equivalent to the original time series. The second moment operator (i.e. the variance) has been used to generate the coarse-grained time series for GSE computation [Costa and Goldberger, 2015]. Thus, at scale 1, the only difference between GSE and SE is the threshold parameter (for more details, see section 4.2.2). Following [Costa et al., 2005] and considering the length of the used windows (4 minutes), we computed the coarse-grained series up to scale 6.

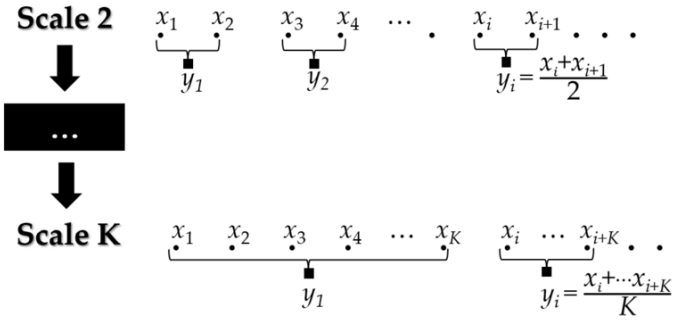


Figure 4.2. Example of the coarse-grained procedure from scale 2 to scale K. [Frassineti et al., 2021d].

For HRV analysis, a 5-minute window should be preferred [Camm et al., 1996]. However, as stated in [Statello et al., 2018, Lucchini et al., 2016, Doyle et al., 2010], a standardized time duration is missing for newborns. Due to the higher newborn heart rate with respect to that of adults, shorter windows are suitable. In this work, a 4-minute window is used as a compromise between a consistent estimation of multiscale HRV entropy indexes [Statello et al., 2018] and the capability to detect seizure events [Doyle et al., 2010]. Thus, considering the average newborn heart rate at rest (100-200 bpm [Doyle et al., 2010]) and the defined window length, we performed computations up to scale 6 to have at least $10^{\wedge m}$ points at each scale and for all the entropy indexes considered [Richman and Moorman, 2000, Costa et al., 2005], where m is the embedding dimension. In this work, we use m=2 for all the entropy indexes. This choice could avoid an inaccurate estimation of entropy values due to a coarse-grained scale at higher scales where the number of points becomes too low [Costa et al., 2005]. Subsections 4.2.2 to 4.2.4 describe the following entropy indexes: AE, SE, GSE, FE and DE. PE has been already introduced in Section 1.3.2. For all the PE scales, also the embedding dimensions m=3 was initially considered [Frassineti et al. 2019]. The multiscale version of PE (MPE) was already evaluated in previous ECG studies [Liu et al., 2017], showing its better capability to enhance differences between groups (pathological vs controls) than the single scale case.

4.2.2 Approximate Entropy, Sample Entropy and Generalized Sample Entropy

Approximate Entropy (AE) [Pincus, 1991] represents the conditional probability that time series that are similar to each other for m consecutive samples (i.e. the embedding dimension) will be similar to each other when one or more samples are known. To introduce the definition of AE, let $\{X_i\}=\{x_1 \ldots x_N\}$ be a time series where N is the number of samples. Moreover, $u_m(i)$ and $u_m(j)$ are vectors of length m (*e.g.* $u_m(i)=\{x_i,\ldots,x_{i+m-1}\}$ for any i and j). Let $n_i^m(r)$ be the number of vectors $u_m(i)$ and $u_m(j)$ that satisfy $d[u_m(i),u_m(j)] \leq r$, where d is the ∞-norm and r a threshold parameter. Therefore $C_i^m(r)=\frac{n_i^m(r)}{N-m+1}$ represents the probability that a vector $u_m(i)$ is close to vector $u_m(j)$. The average of the C_i^m is the probability that any two vectors are within the threshold parameter r. Thus, AE can be defined as in Equation 4.2:

$$AE(m,r,N) = \frac{1}{N-m+1}\sum_{i=1}^{N-m+1} \ln C_i^m(r) - \frac{1}{N-m}\sum_{i=1}^{N-m} \ln C_i^{m+1}(r) \qquad (4.2)$$

In general, low AE values are related to more predictable and regular time series, thus providing a degree of complexity and irregularity [Pincus, 1991]. Richman et al. [Richman and Moorman, 2000] proposed the Sample Entropy (SE) as a modification of AE. The main difference is that SE excludes the self-matches (*i.e.* $i \neq j$), thus reducing the bias of AE. Moreover, SE was proved to be less dependent on time series length, with a higher consistency in different contexts. To date, it is one of the entropy measures most applied to physiological signals analysis [Humeau-Heurtier, 2020].

Recently, Costa et al. [Costa and Goldberger, 2015] defined the Generalized Sample Entropy (GSE) for multiscale experiments. The main difference between GSE and MSE concerns the definition of the coarse-grained time series: instead of using the mean operator, in [Costa and Goldberger, 2015], they used the variance (VAR). GSE can quantify the dynamical time series volatility properties at different scales. Concerning HRV analysis, intermittency in energy and information flows caused by abnormal GSE trends may be related to some pathophysiological processes during both cardiac cycles of activation and recovery (e.g. during depolarization and re-polarization) [Costa and Goldberger, 2015].

For all these entropy measures, and for all the scales, we used as embedding dimension m=2 as in [Richman and Moorman, 2000, Costa et al., 2005], and r=0.2 as threshold parameter of the time series standard deviation for multiscale AE (MAE) and MSE, while for GSE we used r=0.05. This choice allows taking into account the different values in the amplitude of the VAR [Costa and Goldberger, 2015]. The results are reported in Section 4.3.

4.2.3 Fuzzy Entropy

Fuzzy Entropy (FE) can be considered as an extension of SE [Chen et al., 2007]: for a time series $\{X_i\}$ let $u_m(i)$ and $u_m(j)$ be vectors of length m, where m is the embedding dimension. Let $\bar{u}_m(i)$ and $\bar{u}_m(j)$ be the mean values of vectors $u_m(i)$ and $u_m(j)$. The vectors distance $d_{i,j}^m = \max\{|(u_m(i+k)-\bar{u}_m(i))-(u_m(j+k)-\bar{u}_m(j))|,\ 0 \leq k \leq m-1\}$ gives the similarity degree between $u_m(j)$ to $u_m(i)$ that was added through the fuzzy function $D_{i,j}^m(n,r)=\mu\left(d_{i,j}^m,n,r\right)=\exp(-(d_{i,j}^m)^n/r)$. Here r is the threshold parameter and n the exponent parameter. Thus, the FE can be defined by Equation 4.3:

$$FE(m,n,r,N) = \ln\left(\frac{1}{N-m}\sum_{i=1}^{N-m}\left(\frac{1}{N-m-1}\sum_{j=1,j\neq i}^{N-m} D_{i,j}^m(n,r)\right)\right) - \ln\left(\frac{1}{N-m}\sum_{i=1}^{N-m}\left(\frac{1}{N-m-1}\sum_{j=1,j\neq i}^{N-m} D_{i,j}^{m+1}(n,r)\right)\right) \quad (4.3)$$

FE introduces a sort of intermediate state proper of the fuzzy theory, assuming a "*membership degree*" for each point of the time series rather than a conventional two-state classifier as AE and SE [Chen et al., 2007]. Fuzzy Entropy and its multiscale version (MFE) found application in several studies concerning the HRV analysis of pathological subjects [Borin et al., 2021, Humeau-Heurtier, 2020]. As in [Chen et al., 2007], for all the scales an embedding dimension m=2, r=0.2 (of the time series standard deviation) and n=2 were considered.

4.2.4 Distribution Entropy

Distribution Entropy (DE) was proposed by Li et al. [Li et al., 2014]. One of the main advantages of DE with respect to AE and SE, is its higher consistency in the complexity evaluation even for short-term RR signals. Similarly to SE, m is the embedding dimension and $\{X_i\}=\{x_1 \ldots x_N\}$ the original time series, where N is the number of samples. The distance $d_{i,j}^m = \max\{|u(i+k)-u(j+k)|,\ 0 \leq k \leq m-1\}$ among vectors u_i and u_j was computed.

The $d_{i,j}^m$ are the entries of the distance matrix $\boldsymbol{D}=\{d_{i,j}^m\}$. To calculate the distribution entropy, the empirical probability density function (ePDF) of $\boldsymbol{D}$ was estimated by the histogram approach, defining a priori the number of bins B and excluding the self-matches (*i.e.* $i=j$). Thus, the definition of normalized Distribution Entropy using the formula of the Shannon Entropy [Li et al., 2014] is shown in Equation 4.4:

$$DE(B) = -\frac{1}{\log_2 B}\sum_{t=1}^{B} p_t \log_2 p_t \quad (4.4)$$

Where p_t, t =1…B, is the probability of each bin (in this case B=512 [Li et al., 2014]). The multiscale version (MDE) of DE was evaluated in previous studies on ECG [Lee et al., 2018], showing less dependence on the time series length than MSE and MPE.

4.2.5 Statistical analysis

This study aims at evaluating if HRV-entropy indexes allow discriminating between windows with seizure events and seizure-free ones. Specifically, it aims at discriminating between a newborn with seizures and a seizure-free one. To evaluate general differences between the two populations, two statistical analyses were implemented.

The first analysis evaluated if a newborn with seizures might have different characteristics from a seizure-free patient, both for its ictal and interictal activities. Then, it has been evaluated if some differences could be found between seizure windows and interictal ones, that might be helpful in the process of seizure detection [Greene et al., 2007b].

The second analysis was performed to evaluate if a patient with seizures might have different characteristics from a seizure-free patient without distinguishing between seizure windows and interictal windows. In other words, it has been assessed whether the differences in HRV analysis are also found in the interictal activity and not only during seizures. This would support the hypothesis that neonatal seizures alter the cardio-regulatory system of the newborn not only during the ictal events, thus allowing the a priori discrimination between the two groups [Statello et al., 2018]. These two different statistical analyses have been performed for all the multiscale entropy indexes described above. In the following paragraphs, we describe the analyses, the tests used and the assumptions made for each of them.

The hypothesis of normality distribution has been checked through the Shapiro-Wilk test (SW, level of significance $\alpha = 0.05$). As the normality hypothesis was not confirmed for some indexes (MPE, MDE) and some scales for the other indexes, nonparametric tests were applied. All tests have been performed between the 33 median observations of patients with seizure events (here we refer to class "1" for the windows with seizures and to class "int" for the interictal windows) and the 19 median observations of the seizure-free patients (here we refer to the seizure-free windows as class "0"). The first analysis evaluated if differences among various combinations of the three classes ("0"—"1"—"int") exist. To this aim, a multiple comparison test (MCT) was defined, applying a Bonferroni multiple comparison post hoc correction.

We used MCT information obtained by the Kruskal-Wallis test (significance level $\alpha = 0.05$) between the median of the windows of all the patients, considering the test for the three classes "0", "1" and "int" and related pairwise comparisons. We performed this test to evaluate if differences could be found simultaneously between windows with seizure events, seizure-free windows and interictal windows. In particular, the comparison between class "1" and class "0" evaluated the capability of entropy indexes to catch differences between a window with seizure events, i.e., a window from a pathological patient and a window of seizure-free patients. In Section 4.3, the results of the pairwise comparisons referring to each test are reported. Specifically: "0 vs. 1" KW-Test, "int vs. 1" KW-Test, "0 vs. int" KW-Test.

For the second analysis, a Mann-Whitney test was performed (significance level $\alpha = 0.05$) to evaluate if entropy indexes could discriminate between patients with seizure events and seizure-free ones. Here the median of all windows (labelled as "1" and "int") of the newborns with seizures against the median of windows of seizure-free newborns was considered. The statistical results of all the entropy indexes considered for each scale (from 1 to 6) are shown in Section 4.3. We referred to this test as MW-Test.

4.3 Results

In this section, all the statistical results obtained for each multiscale entropy index considered are reported. For each index, a figure shows the multiscale entropy trends for the three classes considered, representing: the seizure-free trend (class "0"), the seizure trend (class "1") and the interictal trend (class "int"). All tests were performed on 33 patients with seizures vs. 19 seizure-free patients. We did not find any statistically significant differences in the comparison "0 vs. int" with the Kruskal Wallis Test for all the entropy indexes; therefore, we did not report them in the related tables. Instead, differences were found for the remaining comparisons: "0 vs. 1" and "int vs. 1" for some multiscale entropy indexes, thus the results are reported. All the KW-Test's p-values were adjusted applying the Bonferroni correction. Table 4.1 reports all the statistical analyses performed using MAE with embedding dimension m=2 and a scale factor from 1 to 6 for multiscale analysis. Moreover, the statistics (median and IQR) for the three groups considered (0, 1, int) are included. In Figure 4.3a the MAE's trends as a function of the scale factor are shown.

Table 4.1. Multiscale Approximate Entropy - Results of the test described in section 2.6. The median and IQR values are reported for the three groups considered (0, 1 and int). Star (*) denotes statistically significant results. [Frassineti et al., 2021d].

MAE Scale factor	0 vs 1 KW-Test p-value	int vs 1 KW-Test p-value	MW-Test p-value	STATS 0 MEDIAN (IQR)	STATS 1 MEDIAN (IQR)	STATS INT MEDIAN (IQR)
1	0.4756	0.1252	0.8048	1.01 (0.82-1.07)	0.92 (0.66-1.04)	1.01 (0.77-1.16)
2	0.0809	0.2318	0.2312	0.88 (0.83-0.97)	0.83 (0.72-0.90)	0.89 (0.77-0.95)
3	0.0129*	0.2160	0.0549	0.83 (0.77-0.88)	0.73 (0.67-0.81)	0.80 (0.73-0.84)
4	0.0937	0.7387	0.2872	0.74 (0.69-0.78)	0.68 (0.65-0.74)	0.71 (0.66-0.77)
5	0.7395	1	0.5686	0.67 (0.64-0.71)	0.66 (0.58-0.69)	0.64 (0.62-0.71)
6	1	1	0.6620	0.61 (0.56-0.64)	0.59 (0.56-0.65)	0.59 (0.56-0.64)

Table 4.2 reports the statistical analysis performed using MSE with embedding dimension m=2 and a scale factor from 1 to 6 for multiscale analysis. The statistics (median and IQR) for the three classes considered are shown. In Figure 4.3b, the MSE trends as a function of the scale factor are shown. As a significant p-value was obtained in the "*int vs. 1*" KW-Test, Figure 4.4 shows the boxplot of the three classes at scale 5.

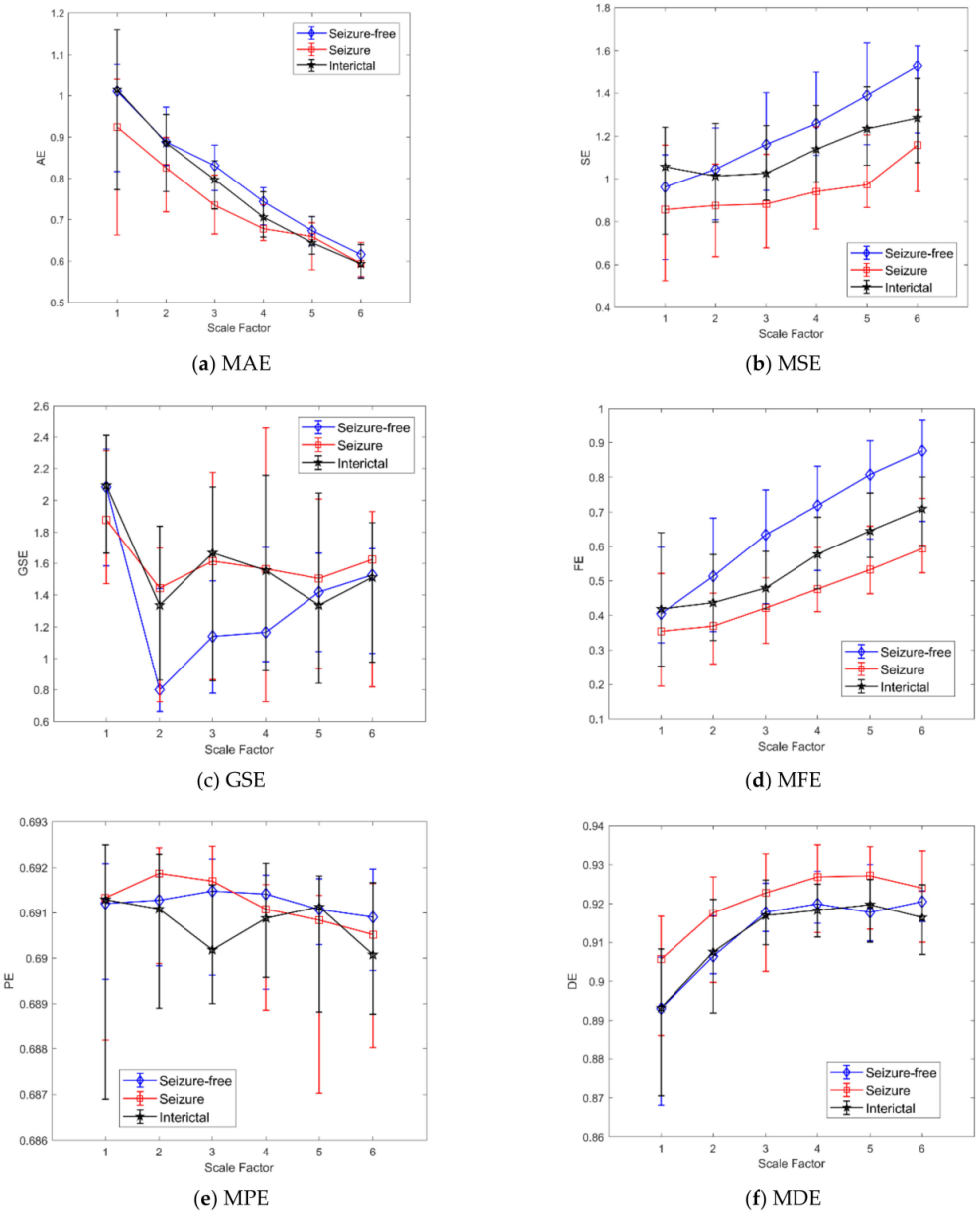
Table 4.2 Multiscale Sample Entropy - Results of the test described in section 2.6. Median and IQR values are reported for the three groups considered (0, 1 and int). Star (*) denotes statistically significant results. [Frassineti et al., 2021d].

MSE Scale factor	0 vs 1 KW-Test p-value	int vs 1 KW-Test p-value	MW-Test p-value	STATS 0 MEDIAN (IQR)	STATS 1 MEDIAN (IQR)	STATS INT MEDIAN (IQR)
1	1	0.1742	0.6213	0.96 (0.62 - 1.11)	0.85 (0.52 - 1.15)	1.05 (0.74 - 1.24)
2	0.1642	0.1733	0.3420	1.04 (0.80 - 1.23)	0.87 (0.63 - 1.06)	1.01 (0.79 - 1.25)
3	0.0054*	0.0786	0.0503	1.16 (0.94 - 1.40)	0.88 (0.67 - 1.11)	1.02 (0.90 - 1.24)
4	0.0022*	0.0990	0.0275*	1.25 (1.11 - 1.49)	0.94 (0.76 - 1.24)	1.13 (0.98 - 1.34)
5	0.0022*	0.0455*	0.0318*	1.38 (1.16 - 1.63)	0.97 (0.86 - 1.20)	1.23 (1.06 - 1.42)
6	0.0064*	0.2387	0.0440*	1.52 (1.21 - 1.62)	1.15 (0.94 - 1.32)	1.28 (1.07 - 1.46)

Table 4.3 reports the statistical analysis performed using GSE with embedding dimension m=2 and a scale factor from 1 to 6 for multiscale analysis. At scale 1, the values are different from the MSE ones because the threshold factor r was set equal to 0.05 [Costa and Goldberger, 2015]. The statistics (median and IQR) for the three classes considered are included. In Figure 4.3c, the GSE trends as a function of the scale factor are shown.

Table 4.3. Generalized Sample Entropy - Statistical results of the test described in section 2.6. The descriptive statistics (median and IQR) related to the three groups considered (0, 1 and int) are reported. Star (*) denotes statistically significant results. [Frassineti et al., 2021d].

GSE Scale factor	0 vs 1 KW-Test p-value	int vs 1 KW-Test p-value	MW-Test p-value	STATS 0 MEDIAN (IQR)	STATS 1 MEDIAN (IQR)	STATS INT MEDIAN (IQR)
1	1	0.4084	0.9092	2.08 (1.58 - 2.32)	1.87 (1.47 - 2.31)	2.09 (1.66 - 2.41)
2	0.5154	1.0000	0.0574	0.80 (0.66 - 1.44)	1.44 (0.72 - 1.69)	1.33 (0.86 - 1.83)
3	0.2617	1.0000	0.0166*	1.14 (0.77 - 1.49)	1.61 (0.86 - 2.17)	1.66 (0.85 - 2.08)
4	0.9657	1.0000	0.0681	1.16 (0.97 - 1.70)	1.56 (0.72 - 2.45)	1.55 (0.92 - 2.15)
5	1	1.0000	0.0908	1.41 (1.04 - 1.66)	1.50 (0.93 - 2.00)	1.33 (0.84 - 2.04)
6	1	1.0000	0.1434	1.52 (1.03 - 1.69)	1.62 (0.82 - 1.93)	1.51 (0.97 - 1.85)



(**a**) MAE (**b**) MSE

(c) GSE (**d**) MFE

(**e**) MPE (**f**) MDE

Figure 4.3. Comparison of the proposed approaches results for scales 1 to 6. (a) Approximate Multiscale Entropy.(b) Multiscale Sample Entropy. (c) Generalized Multiscale Sample Entropy. (d) Multiscale Fuzzy Entropy. (e) Multiscale Permutation Entropy. (f) Multiscale Distribution Entropy. The trends related to the seizure-free windows (◊ marker), seizure windows (□ marker) and interictal windows (☆ marker) are shown. The values at each scale represent the median and iqrs among patients. [Frassineti et al., 2021d].

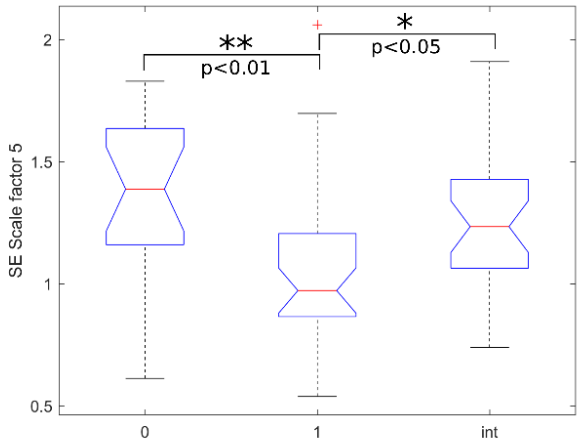


Figure 4.4. SE boxplots of the groups (0-1-int) with scale factor 5, where * and ** respectively denote statistically significant and highly statistically significant results obtained with the MCT test. [Frassineti et al., 2021d].

In Table 4.4 the statistical analysis performed using FE is reported, with embedding dimension m=2 and a scale factor from 1 to 6 for multiscale analysis. As for MSE, the statistics (median and IQR) for the three classes considered are reported. In Figure 4.3d, the MFE trends as a function of the scale factor are shown. As a significant p-value in the "*int vs. 1*" KW-Test was obtained, Figure 4.5 shows the boxplot of the three classes at scale 5 and the corresponding multi-comparison analysis.

Table 4.4. Multiscale Fuzzy Entropy - Statistical results of the test described in section 2.6. The descriptive statistics (median and IQR) related to the three groups considered (0, 1 and int) are reported. Star (*) denotes statistically significant results. [Frassineti et al., 2021d].

MFE Scale factor	0 vs 1 KW-Test p-value	int vs 1 KW-Test p-value	MW-Test p-value	STATS 0 MEDIAN (IQR)	STATS 1 MEDIAN (IQR)	STATS INT MEDIAN (IQR)
1	0.9564	0.2411	0.8344	0.40 (0.32 - 0.59)	0.35 (0.19 - 0.52)	0.41 (0.25 - 0.64)
2	0.0316*	0.3318	0.0526	0.51 (0.35 - 0.68)	0.36 (0.25 - 0.46)	0.43 (0.32 - 0.57)
3	0.0052*	0.2207	0.0150*	0.63 (0.43 - 0.76)	0.42 (0.31 - 0.50)	0.47 (0.41 - 0.58)
4	0.0031*	0.1444	0.0204*	0.71 (0.53 - 0.83)	0.47 (0.41 - 0.59)	0.57 (0.47 - 0.68)
5	0.0013*	0.0469*	0.0194*	0.80 (0.62 - 0.90)	0.53 (0.46 - 0.65)	0.64 (0.56 - 0.75)
6	0.0011*	0.0864	0.0083*	0.87 (0.67 - 0.96)	0.59 (0.52 - 0.73)	0.70 (0.59 - 0.80)

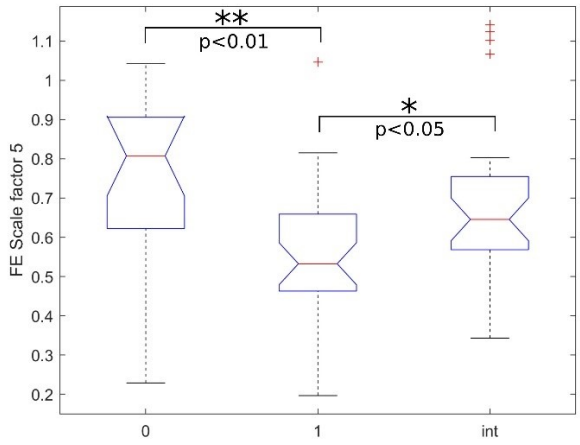


Figure 4.5. FE boxplots of the groups (0-1-int) with scale factor 5, where * and ** respectively denote statistically significant and highly statistically significant results obtained with the MCT test. [Frassineti et al., 2021d].

Table 4.5 reports the statistical analysis performed using PE with embedding dimension *m*=2 and a scale factor from 1 to 6 for multiscale analysis. Moreover, the statistics (median and IQR) for the three classes considered are reported. In Figure 4.3e, the MPE trends as a function of the scale factor are shown.

Table 4.5. Multiscale Permutation Entropy - Statistical results of the test described in section 2.6. All tests were performed on 33 patients with seizures vs. 19 seizure-free patients. The descriptive statistics (median and IQR) related to the three groups considered (0, 1 and int) are reported. Star (*) denotes statistically significant results. [Frassineti et al., 2021d].

MPE Scale factor	0 vs 1 KW-Test p-value	int vs 1 KW-Test p-value	MW-Test p-value	STATS 0 MEDIAN (IQR)	STATS 1 MEDIAN (IQR)	STATS INT MEDIAN (IQR)
1	1	1	0.7323	0.6912 (0.6895 - 0.6921)	0.6913 (0.6882 - 0.6925)	0.6913 (0.6869 - 0.6925)
2	0.5050	0.9519	0.3769	0.6913 (0.6898 - 0.6918)	0.6919 (0.6899 - 0.6924)	0.6911 (0.6889 - 0.6923)
3	1	0.3827	0.7467	0.6915 (0.6896 - 0.6922)	0.6917 (0.6890 - 0.6925)	0.6902 (0.6890 - 0.6916)
4	1	1	0.8792	0.6914 (0.6893 - 0.6918)	0.6911 (0.6889 - 0.6916)	0.6909 (0.6896 - 0.6921)
5	1	1	0.4529	0.6911 (0.6903 - 0.6918)	0.6908 (0.6870 - 0.6914)	0.6911 (0.6888 - 0.6918)
6	1	1	0.4359	0.6909 (0.6897 - 0.6920)	0.6905 (0.6880 - 0.6917)	0.6901 (0.6888 - 0.6917)

Table 4.6 reports the statistical analysis performed using DE with embedding dimension *m*=2, number of bins 512 and a scale factor from 1 to 6 for multiscale analysis. As for MPE, the statistics (median and IQR) for the three classes considered are reported. In Figure 4.3f, the MDE trends as a function of the scale factor are shown.

Table 4.6. Multiscale Distribution Entropy - Statistical results of the test described in section 2.6. The descriptive statistics (median and IQR) related to the three groups considered (0, 1 and int) are reported. Star (*) denotes statistically significant results. [Frassineti et al., 2021d].

MDE Scale factor	0 vs 1 KW-Test p-value	int vs 1 KW-Test p-value	MW-Test p-value	STATS 0 MEDIAN (IQR)	STATS 1 MEDIAN (IQR)	STATS INT MEDIAN (IQR)
1	0.3664	0.2952	0.2312	0.8931 (0.8681 - 0.9062)	0.9057 (0.8859 - 0.9167)	0.8932 (0.8706 - 0.9083)
2	0.8091	0.5365	0.7323	0.9064 (0.9019 - 0.9168)	0.9176 (0.8998 - 0.9269)	0.9075 (0.8919 - 0.9211)
3	1	0.6727	1.0000	0.9178 (0.9128 - 0.9253)	0.9228 (0.9026 - 0.9328)	0.9169 (0.9094 - 0.9261)
4	1	0.2114	0.7756	0.9200 (0.9149 - 0.9283)	0.9269 (0.9125 - 0.9351)	0.9183 (0.9114 - 0.9250)
5	1	0.7985	0.8942	0.9177 (0.9104 - 0.9300)	0.9272 (0.9134 - 0.9347)	0.9198 (0.9100 - 0.9263)
6	1	0.3935	0.9697	0.9205 (0.9154 - 0.9233)	0.9240 (0.9100 - 0.9336)	0.9164 (0.9069 - 0.9249)

In Table 4.7 we show the descriptive statistics (median and IQR) concerning the complexity index (CI) and the relative slope [Costa et al., 2005] of all the multiscale entropy indexes and the classes considered.

Table 4.7. Descriptive statistics (median and IQR) for all the entropy indexes and the classes considered ("0-1-int") concerning the complexity index (CI) with the corresponding slope (+1 if positive -1 if negative). Stars (*) denote significant differences between class "0" and "1" (Mann-Whitney Test p-value < 0.05, after Bonferroni correction). [Frassineti et al., 2021d].

Multiscale Entropy Index	CI - median (iqr) class "0"	CI - median (iqr) class "1"	CI - median (iqr) class "INT"
MAE	-3.9 (-4.2 : -3.7)	-3.6 (-3.9 : -3.1)	-3.8 (-4.0 : -3.4)
MSE*	6.3 (5.1 : 7.2)	4.2 (2.6 : 5.6)	5.0 (3.3 : 6.2)
MFE*	3.1 (2.3 : 3.7)	2.3 (1.7 : 2.7)	2.6 (1.6 : 3.1)
GSE	-5.9 (-8.1 : -3.1)	-2.1 (-7.0 : 9.4)	-5.7 (-8.2 : -1.3)
MPE	-3.4 (-3.5 : 3.5)	-3.4 (-3.5 : 3.4)	-3.4 (-3.5 : 3.4)
MDE	4.6 (4.5 : 4.6)	4.6 (4.0 : 4.6)	4.6 (4.5 : 4.6)

4.4 Discussion

This work applies multiscale entropy analysis to HRV dynamics with the aim of providing useful information about possible autonomic nervous system dysregulation occurring during seizures in newborns.

Table 4.1 gives some significant differences between groups with MAE. However, the multiscale analysis shows the well-known limits of this metric in the analysis of short time series [Richman and Moorman, 2000]: at higher scales, it was no longer possible to resolve differences (from scale 4 in Table 1, "0 vs. 1" KW-Test). Moreover, the MW-Test for all the scales considered shows that MAE might be not suitable to discriminate between a patient with seizures and a seizure-free one. The results reported in Tables 4.2 and 4.4 show that SE and FE are capable of discriminating between a patient with seizure events from a seizure-free one ("0 vs. 1" KW-Test and MW-Test). The multiscale analysis proved to be crucial to detect these differences. In fact, without any scale factor (i.e., $s = 1$, Equation (4.1)), no differences were found among subjects. Instead, Tables 4.2 and 4.4 show that windows with seizure events have lower values of SE and FE than those obtained for the interictal and seizure-free windows. This is partially in line with what was already found for childhood seizures using entropy indexes in EEG analysis [Frassineti et al., 2019]. Furthermore, the results shown in Table 4.7 summarize what was found with the statistical analysis. As an example, the CI for MSE and MFE on windows with seizure events are on average lower than those in seizure-free windows (Mann-Whitney test, p-values < 0.05 with Bonferroni correction). Instead, for MPE or MAE the CI values are very similar among the classes (Mann-Whitney test, p-values > 0.05 with Bonferroni correction).

Thus, considering the results obtained in the "0 vs. 1" KW-Test and MW-Test for MSE and MFE (Tables 4.2 and 4.4), it seems that differences between a newborn with seizures and a seizure-free one do not exist only during the ictal events but also during the whole interictal activity. These results seem to confirm that neonatal seizure events may produce a direct or indirect continuous alteration at the level of the cardio-regulatory system. Multiscale entropy indexes may detect these abnormal heart rate dynamics, probably connected to a reduced variability or transient decelerations in heart rate dynamics [Lake et al., 2002].

Moreover, a slight difference between the results obtained with FE and SE could indicate a higher accuracy of FE. Results shown in Tables 4.2 and 4.4 highlighted that the MFE index was able to catch the differences on more scales than MSE, showing more consistency and intrinsic robustness against noise in the time series [Borin et al., 2021]. The MSE and MFE entropies showed promising results, thus suggesting that higher scale levels should be exploited using variants of MSE, such as Modified multiscale entropy or others [Humeau-Heurtier, 2015].

GSE (Table 4.3) deserves a different discussion from MAE. The significant differences in volatility found by the MW-Test at scale factor 3 suggest a different behaviour between groups about instantaneous variability of HRV properties. As suggested by Costa et al. [Costa and Goldberger, 2015], this behaviour might be related to possible abnormal heart dynamics during the cardiac cycle of activation and recovery (e.g., during depolarization and repolarization). However, the relationship between these findings and physiological dynamics is still an open issue, and further analysis is required. The obtained results suggest that a coarse-grained procedure based on the mean value should be preferred over variance-based methods. However, as stated by Costa et al. [Costa and Goldberger, 2015], further analysis about GSE should be made by increasing the window size or using different entropy variants.

Concerning these findings, in Figure 6 the multiscale trends obtained for MW-Test (from Tables 4.2–4.4) are reported. The Figure shows that MFE, MSE and GSE were able to catch differences between seizures and seizure-free patients. Specifically, Figure 4.6 shows the cumulative trends related to patients with seizures for all the windows extracted, both those with seizure events (class "1") and the interictal ones (class "int"), showing that differences still exist between the two groups. All the trends, starting from scale 2 for MFE and scale 3 for MSE, show that differences between a newborn with seizure and a seizure-free one can be found during or close to seizure events and during the interictal periods.

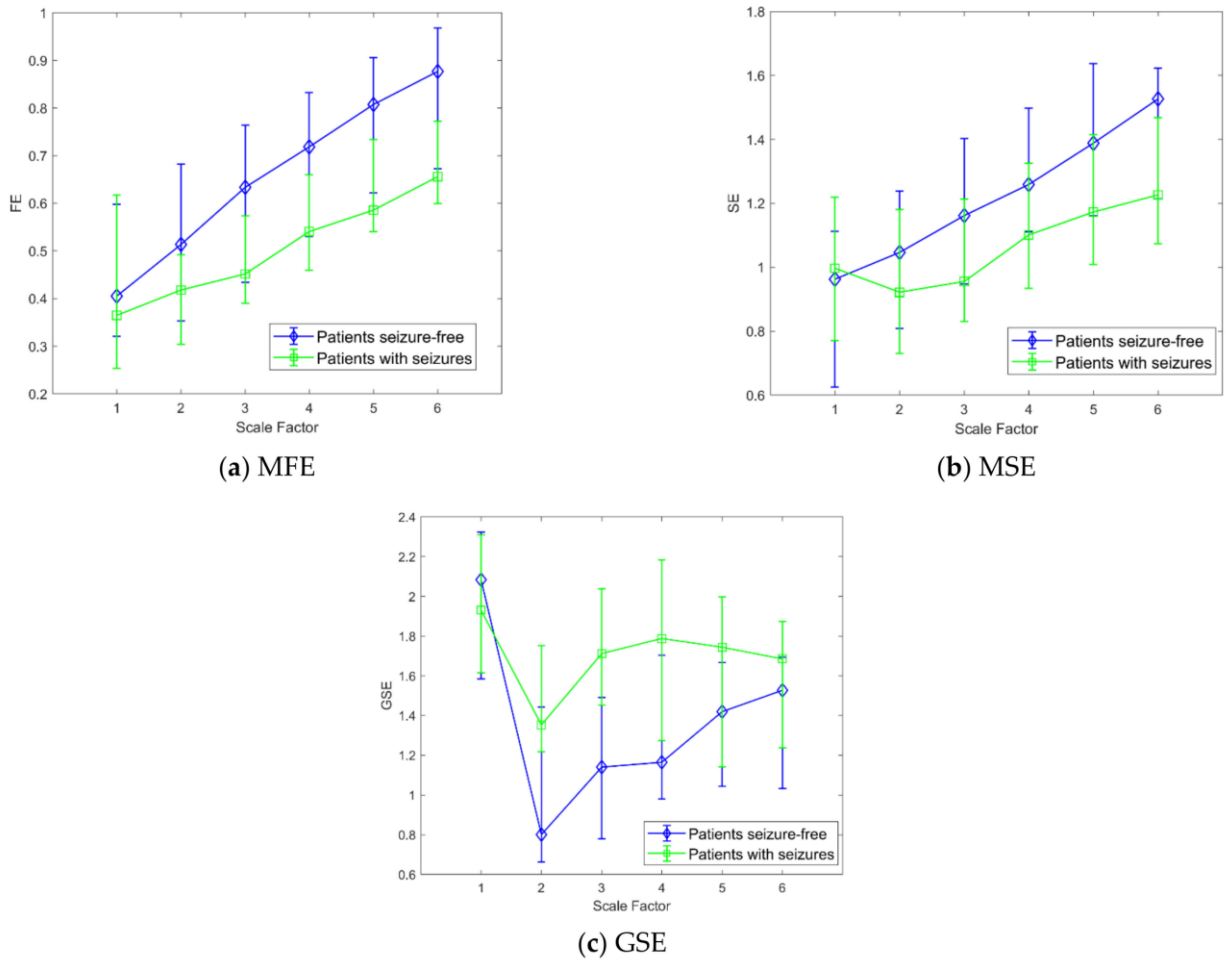


Figure 4.6. MFE (a), MSE (b), and GSE (c) trends considering all the windows for the patients with seizures (□ markers) compared to the seizure-free patients (◊ markers). The values at each scale represent the median and IQR among patients. [Frassineti et al., 2021d].

Instead, as shown in Tables 4.1, 4.5, and 4.6 for MAE, MPE, and MDE, we did not find any significant differences among groups. This suggests that, according to the used embedding dimension and threshold parameter, these indexes might not be helpful for HRV multiscale entropy analysis in newborns with seizures. As shown in the Results Section, differences between the methods considered here exist. Although an exhaustive answer about the possible reasons for all the entropies is challenging, it could be argued that the difference lies in the basic properties and limits of each method previously described. Limits of MAE and MSE have been discussed in [Humeau-Heurtier, 2015] and [Humeau-Heurtier, 2020]. Moreover, limits of MPE were found when applied to short time series, or with a low signal-to-noise ratio [Porta et al, 2015]. The present work focuses on a limited

set of parameters and indexes and cannot be generalized, anyway it shows that the choice of the proper entropy index is crucial and requires a deeper analysis for a consistent and robust investigation.

Finally, although "0 vs. 1" KW-Test and MW-Test confirmed the differences between patients with seizures and seizure-free patients, we found that only MSE and MFE at scale factor 5 could discriminate between interictal windows and windows with seizure events (Tables 4.2 and 4.4, "int vs. 1" KW-Test: p-value < 0.05). The result suggests that multiscale entropy indexes might not be suitable for an intra-patient analysis, an essential prerequisite for their implementation in a patient-independent ECG-based NSD [Frassineti et al., 2021c, Greene et al., 2007b]. However, this goal was out of the aims of this work and will be the subject of future studies.

In summary, our results confirm that HRV multiscale entropy analysis may provide helpful information for the characterization of neonatal seizures. As shown in [Bersani et al, 2021], HRV analysis can provide a reliable marker for HIE, one of the most common neonatal seizures aetiologies.

Nevertheless, our analysis has some limitations: in general, the Helsinki Dataset includes a large number of newborns with episodes of asphyxia/HIE, that may have partly biased the implemented methods in differentiating between the groups. Moreover, our analysis did not cover extreme events such as sudden infant death syndrome (SIDS) that could be associated with an abnormal cardiac activity [Buchanan, 2019]. Thus, further studies are required to extend our findings to all the possible aetiologies behind neonatal seizures. To the best of our knowledge, the Helsinki Dataset is the most extensive public dataset for neonatal seizures in terms of the number of patients [Olmi et al., 2021], but we tested our methods on this dataset only. Thus, our findings may be considered preliminary and they need to be confirmed after their application to other datasets. This work does not consider all the entropy metrics proposed in the literature, thus the optimal entropy measures for this task will be the subject of further research. About MSE analysis, several variants were recently proposed [Humeau-Heurtier, 2015] and will be considered as future developments of the methods, especially for the analysis of short time series. Promising methods, even for short time series, such as AvgApEn, AvgSampEn proposed by C. Karmakar et al. [Karmakar et al., 2020], ipApEn and ipSampEn introduced by G. Valenza et al. [Valenza et al, 2015] may be valid alternatives to be evaluated.

Another issue concerns the choice of the most suitable window length. In this work, we used windows of 4 min of duration to detect autonomic variations in newborns [Statello et al., 2018]. This choice allowed a consistent multiscale analysis and the discrimination between windows with seizure events and interictal windows. It is worthwhile noting that the choice of 4-min windows may not be the best and it might depend on the specific dataset. Thus, further studies would be focused on different datasets to confirm this finding. However, to detect the onset and the offset of an ictal event by HRV analysis, e.g., for a real-time evaluation, windows with shorter duration should be considered in future studies. Our results concern a single embedding dimension and a single threshold parameter. Although our settings are confirmed in the literature [Humeau-Heurtier, 2020], an exhaustive research involving all the possible combinations of parameters, indexes and scales could be exploited. Moreover, a deeper analysis about parameters for multiscale entropy indexes could be taken into account in the future, in particular for indexes like PE and DE that were not able to detect the differences between newborns.

HRV entropy indexes seem to be appropriate to describe neonatal seizures with a specific aetiology such as the hypoxic-ischemic encephalopathies or, in general, with asphyxia episodes [Michniewicz et al., 2020, Locatelli et al, 2020]. Moreover, cardio-regulatory system differences between a newborn with seizures and a seizure-free one might also be present during interictal activity rather than only during or close to an ictal event.

4.5. BHIs in newborns

In this chapter a first analysis regarding the use of multiscale HRV entropy indexes as a support for neonatal seizure detection and characterization was provided. Considering the promising results, they suggest that an interaction between ANS and Central Autonomic Nervous System (CNS) during neonatal seizures might exist. It is well known that a mutual exchange of information exists between the cortical activity of the CNS and the ANS. For example, cardiac activity can be altered by inputs from baroreceptors, chemoreceptors, and other sources [Silvani et al., 2016]. Thus, the analysis and the characterization of the brain-heart interactions (BHIs) during physiological and pathological events is of great clinical interest [Silvani et al., 2016]. The ANS-to-CNS system interaction was investigated and modelled using physiological signals such as the electroencephalogram (EEG) and the heart rate variability (HRV) [Schiecke et al., 2019]. Moreover, several methods were proposed to measure or model these interactions, such as Granger causality, Transfer entropy and the convergent cross mapping (CCM) [Schiecke et al., 2019]. These methods have been applied in various neuroscience fields such as polysomnography [Faes et al., 2015], stress assessment, mood disorders, emotion recognition [Valenza et al., 2016] and epilepsy [Schiecke et al., 2016].

In particular, the study of BHI in epilepsy could have relevant diagnostic and therapeutical applications in epilepsy to detect signs or symptoms related to a sudden unexpected death in epilepsy (SUDEP) [Costagliola et al., 2021]. The analysis of interactions between physiological systems has already been performed for newborns, mainly to assess neurovascular coupling [Hendrikx et al., 2019]. However, to the best of our knowledge, a specific investigation of BHIs in newborns with seizure events using EEG and HRV signals is still scarce [Hendrikx et al., 2019]. If confirmed this information could be used as support to ECG/HRV-based as well as the already explained multiscale HRV entropy measures. For these reasons, in this final part of the Chapter 4 a first investigation about BHIs in newborns with and without seizures using EEG and HRV signals is presented. The aim is to assess if different behaviors may exist between the CNS and ANS for the two populations considered. We used the CCM method [Sugihara et al., 2012], already applied to the analysis of BHI in specific childhood epilepsy [Schiecke et al., 2016]. To the best of our knowledge, the study presented here is one of the first studies applying CCM methods to characterize BHIs during neonatal seizures.

4.5.1 CCM methods

To obtain a uniform comparison with multiscale HRV entropy results, all the proposed methods were implemented using the same cohort: 33 newborns with seizure events and 19 seizure-free from the Helsinki Dataset [Stevenson et al., 2019]. Regarding the EEG derivations, as in [Frassineti et al., 2020], the same bipolar configuration was used: F4-C4, C4-O2, F3-C3, C3-O1, T4-C4, C4-Cz, Cz-C3 and C3-T3. All the EEG signals were band-pass FIR filtered in the bandwidth 0.25-16 Hz. ECG signals were analyzed to extract HRV time series. Both ECG and EEG underwent a sub-windowing procedure of 30s of duration [Schiecke et al., 2016].

To increase the signal-to-noise ratio (SNR), ECGs were first pass-band filtered in the 0.05-45 Hz bandwidth. Then inter-beat-interval (IBI) time series were obtained using the Pan-Tompkins' algorithm [Pan and Tompkins, 1985]. The HRV signals were eventually interpolated to have the same number of samples of the EEG signals. The following interpolation methods were checked: linear, nearest neighbour and French-Holden algorithm [Schiecke et al., 2019]. Both EEG and the interpolated HRV were downsampled to 16 Hz, obtaining 480 samples in each window. For each EEG derivation and each 30s window, we evaluated the CCM correlation coefficients between the

EEG and the HRV signals. The CCM approach is a nonlinear method to assess causality between two time series X and Y, observing the correspondence between the so-called "Shadow Manifolds" M_X and M_Y, built using lagged coordinates from the original time series X and Y [Sugihara et al., 2012]. The lagged versions depend on: the embedding dimension D, the time lag τ and the library length L [Sugihara et al., 2012]. In our case, L=480 i.e. the number of window's samples, $2 \leq D \leq 8$ and $1 \leq \tau \leq 5$ [Schiecke et al., 2016].

The interactions between the two systems were quantified by the CCM correlation, defined as the absolute value of the Pearson correlation coefficient (ρ) between the original time series and an estimation using the CCM with the other time series. Thus, we obtained two CCM indexes (55) and (56) for all the 8 derivations considered, defined as [Sugihara et al., 2012]:

$$CCM_{EEG \to HRV} = |\rho(EEG, EEG/M_{HRV})| \quad (4.5)$$

$$CCM_{HRV \to EEG} = |\rho(HRV, HRV/M_{EEG})| \quad (4.6)$$

Moreover, we computed the Average Degree metrics [Hendrikx et al., 2018] as the mean of CCM values, both for (4.5) and (4.6), between all the derivations considered.

The overall mean of CCM coefficients for each patient was computed, distinguishing between seizure and seizure-free subjects. First, we tested if CCM values (4.5) and (4.6) and the Average Degree's CCM values were statistically different between the two groups. The hypothesis of normality distribution was checked applying the Shapiro-Wilk test (level of significance α=0.05). As this hypothesis was not confirmed, we applied the non-parametric Mann-Whitney test (Test MW, level of significance α=0.05).

Moreover, a surrogate analysis was added to check if the CCM values were due to chance or random fluctuations within the time series or represent an interaction between the two systems. To this aim, for each window, a set of 100 surrogates was built from the HRV interpolated signal using the amplitude-adjusted Fourier transform (AAFT) method [Lancaster et al., 2018]. Then for the surrogate sets, the CCM correlation coefficients (1) and (2) and the Average Degree were computed. A CCM value was significant if it was larger than the significance threshold: $T_{surr}=\mu_{surr}+2\sigma_{surr}$; where μ_{surr} and σ_{surr} are respectively the mean and the standard deviation values of the metrics obtained from the surrogate sets [Perrella et al., 2018]. Statistical differences between surrogate CCM values and CCM values from the original time series were tested applying a non-parametric Mann-Whitney test (level of significance α=0.05). The null hypothesis was the lack of BHIs and that the surrogate pairs destroy the original CCM correlation.

4.5.2 CCM Results

In Table 4.8, the statistical results obtained on $CCM_{EEG \to HRV}$ mean values between 33 patients with seizures and 19 seizure-free ones are shown. The same analysis for the Average Degree parameter is reported in the last row. The descriptive statistics are related to the following CCM's parameters: D=3, τ=1, L=480. The interpolation method selected for this experiment was the linear one. The same statistical analysis related to $CCM_{HRV \to EEG}$ is shown in Table 4.9. Table 4.10 concerns the descriptive statistics of CCM parameters related to surrogate analysis. The average significant thresholds and their standard deviations are shown for patients with seizures and seizure-free ones. Star (*) denotes a statistically significant difference with the CCM values computed to the original time series. A non-parametric Mann-Whitney test (level of significance α=0.05) was applied.

Table 4.8. Results of statistical tests for the parameter $CCM_{EEG \rightarrow HRV}$ performed between each derivation and HRV signal. The descriptive statistics (mean μ ± standard deviation σ) are shown. Star (*) denotes statistically significant results. [Frassineti et al., 2022a].

	CCM EEG → HRV		**Test MW**
Derivation	**Seizure-free patients μ±σ**	**Patients with seizures μ±σ**	**p-value**
F4-C4	0.23 ± 0.08	0.18 ± 0.05	0.010*
C4-O2	0.24 ± 0.07	0.19 ± 0.06	0.001*
F3-C3	0.22 ± 0.10	0.18 ± 0.07	0.106
C3-O1	0.22 ± 0.07	0.19 ± 0.09	0.154
T4-C4	0.24 ± 0.08	0.19 ± 0.06	0.020*
C4-Cz	0.25 ± 0.07	0.19 ± 0.06	0.012*
Cz-C3	0.23 ± 0.08	0.19 ± 0.08	0.018*
C3-T3	0.21 ± 0.10	0.18 ± 0.09	0.223
Average Degree	0.23 ± 0.08	0.20 ± 0.07	0.029*

Table 4.9. Results of statistical tests for $CCM_{HRV \rightarrow EEG}$. The descriptive statistics (mean μ ±standard deviation σ) are shown. Star (*) denotes statistically significant results. [Frassineti et al., 2022a].

	CCM HRV → EEG		**Test MW**
Derivation	**Seizure-free patients μ ± σ**	**Patients with seizures μ ± σ**	**p-value**
F4-C4	0.11 ± 0.03	0.11 ± 0.04	0.530
C4-O2	0.10 ± 0.03	0.10 ± 0.05	0.287
F3-C3	0.12 ± 0.04	0.10 ± 0.04	0.493
C3-O1	0.09 ± 0.03	0.10 ± 0.03	0.044*
T4-C4	0.10 ± 0.02	0.11 ± 0.05	0.071
C4-Cz	0.11 ± 0.02	0.11 ± 0.03	0.196
Cz-C3	0.10 ± 0.03	0.11 ± 0.04	0.371
C3-T3	0.09 ± 0.02	0.10 ± 0.03	0.019*
Average Degree	0.10 ± 0.02	0.11 ± 0.04	0.183

Table 4.10 Descriptive statistics (mean μ ± standard deviation σ) from Surrogate Analysis. Star (*) denotes significant differences between surrogates CCM values and their respective values shown in Table 4.8 and 4.9. [Frassineti et al., 2022a].

	Test Surrogates $T_{surr}=\mu_{surr}+2\sigma_{surr}$			
	CCM EEG → HRV		**CCM HRV → EEG**	
Derivation	**Seizure-free patients T_{surr} (μ ± σ)**	**Patients with seizures T_{surr} (μ ± σ)**	**Seizure-free patients T_{surr} (μ ± σ)**	**Patients with seizures T_{surr} (μ ± σ)**
F4-C4	0.12± 0.02*	0.10± 0.02*	0.10 ± 0.01	0.10± 0.02
C4-O2	0.13± 0.03*	0.11± 0.03*	0.09 ± 0.02	0.10± 0.02
F3-C3	0.12± 0.03*	0.10± 0.02*	0.10 ± 0.02	0.10± 0.01
C3-O1	0.12± 0.03*	0.11± 0.02*	0.09 ± 0.02	0.10± 0.02
T4-C4	0.13± 0.03*	0.12± 0.03*	0.10 ± 0.02	0.10± 0.02
C4-Cz	0.14± 0.03*	0.11± 0.02*	0.10± 0.01	0.10± 0.01*
Cz-C3	0.13± 0.03*	0.11± 0.02*	0.10± 0.02	0.10± 0.02
C3-T3	0.12± 0.03*	0.10± 0.02*	0.09± 0.02	0.10± 0.02
Average Degree	0.12± 0.03*	0.10± 0.02*	0.09± 0.01*	0.10± 0.02

4.5.3 Discussion regarding CCM and BHIs in neonatal seizures

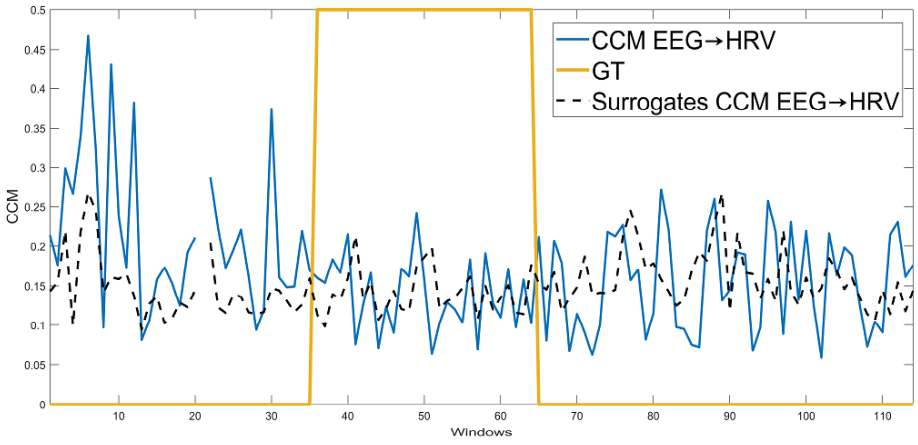
This work aims at evaluating whether CCM analysis could provide helpful and significant information about BHIs in newborns with and without seizures. As shown in Table 4.8, significant results were obtained for $CCM_{EEG \rightarrow HRV}$ values for several derivations (5 out of the 8 considered) and for the Average Degree parameter (p-values<0.05): the interaction between CNS and ANS may differ in patients with seizures and seizure-free ones. Moreover, the average values of $CCM_{EEG \rightarrow HRV}$ were lower for patients with seizures than for seizure-free ones.

This suggests that neonatal seizures might significantly alter the neuronal interplay between the two systems, making the ANS less "predictable" in response to variation of the CNS or cortical activity [Schiecke et al., 2019]. Low values of CCM mean a low causality relationship between two systems [Sugihara et al., 2012]. We did not obtain significant results for all the derivations considered, as neonatal seizures are mainly focal ones [Pressler et al., 2021], thus, some cerebral areas may not be involved during ictal events for most patients in our dataset. Since the surrogate thresholds for $CCM_{EEG \rightarrow HRV}$ were lower than the original values (Table 4.10 and Table 4.8, respectively), these interactions might be due to specific relationships between CNS and ANS and not to random fluctuations of the time series.

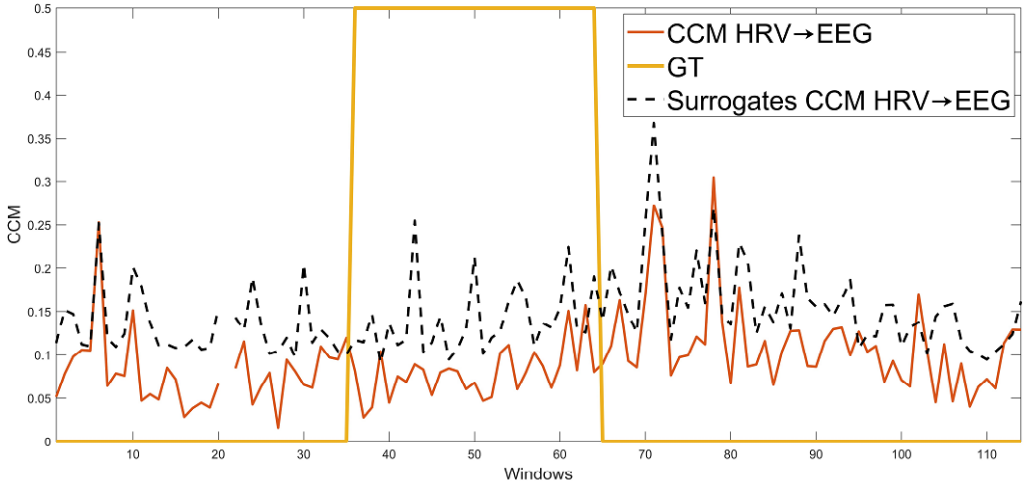
However, the same cannot be said for the $CCM_{HRV \rightarrow EEG}$ values: although we obtained significant results for two derivations (C3-O1 and C3-T3 in Table 4.9), these results must be considered with caution because the surrogate values (Table 4.10) did not show significant differences between them (p-value > 0.05). Thus, these interactions might be due to chance and not to a true relationship between ANS and CNS. Figure 4.7a. and 4.7b show the Average Degree's CCM trends of a single patient and their surrogate analysis. As shown in Figure 4.7a, the surrogate values remain below the $CCM_{EEG \rightarrow HRV}$, while $CCM_{HRV \rightarrow EEG}$ values remain lower than surrogates in almost all windows. Results suggest that in patients with seizure events the heart dynamics could be altered by the CNS activity but not the opposite. This is in agreement with [Statello et al., 2021]: seizures may alter the ANS dynamics and not only the CNS one. The CCM approach seems to catch differences in BHIs between newborns with and without seizures.

However, our analysis has some limits. First, the choice of EEG montage is a critical point that might alter the BHI results. Thus, further analysis is needed to find the best montage for neonatal BHI analysis. Anyway, as shown in Table 4.8, multichannel analysis can provide better information than a single derivation analysis because some derivations may not give significant interactions. Another critical issue concerns the choice of CCM parameters D, τ and L, which could not be the optimal ones if used on other datasets. As an example, varying the downsampling factor might better characterize interactions due to different brain waves. Thus, an exhaustive evaluation should be performed when CCM analysis is applied to other datasets [Schiecke et al., 2016]. This point is outside of the aims of the present research, but could be addressed in future studies. Moreover, in this work the CCM approach was chosen to characterize the BHIs in newborns, however in future studies alternative approaches, such as bivariate and multivariate methods, could be tested [Faes et al., 2016].

This work shows that BHIs might differ in newborns with seizures and seizure-free ones. Future studies could focus on BHI differences among patients with seizures, focusing on interictal and ictal periods [Frassineti et al., 2021d] or between pre-ictal and post-ictal periods [Frassineti et al., 2021a]. Other studies could investigate BHIs among different aetiologies [Olmi et al., 2021] or after pharmacological treatments [Costagliola et al., 2021].



(a)



(b)

Figure 4.7. BHI analysis for a single patient with a single seizure event. (a) Average Degree's $CCM_{EEG \rightarrow HRV}$ trend (solid blue line) and the corresponding surrogate values for each window (dashed line). (b) Average Degree's $CCM_{HRV \rightarrow EEG}$ trend (solid red line) and the corresponding surrogate values for each window (dashed line). The orange line between windows 35 and 65 represents the time occurrence of the seizure event (GT= ground truth). [Frassineti et al., 2022a].

4.6. Conclusions

In this chapter two different analysis regarding the possible involvement of ANS during neonatal seizures and how to quantify it have been presented.

The first analysis concerned the evaluation of multiscale HRV entropy indexes for the characterization of neonatal seizures. Specifically, the capability of entropy measures to detect abnormal heart rate dynamics during or close to ictal events was exploited. Entropy measures were analyzed to discriminate between newborns with seizures and seizure-free ones. We found that Multiscale Sample and Fuzzy Entropy, from the scale factors 3 and 2 respectively, show significant differences between the two groups. Thus, the multiscale approach allows characterizing the ictal events that could not be detected with a single scale approach. Moreover, interictal activity showed significant differences between patients with seizure and seizure-free ones.

Though our results are promising, it is noteworthy that HRV analysis may not be specific enough for neonatal seizure detection. For example, motor activity during seizures could lead to changes in heart rate and its variability, although neonatal seizures are often without a clear motor activity and can

only be recognized from their electroclinical characteristics [Pressler et al., 2021]. Thus, more studies are required to better clarify the neuro-vascular mechanisms occurring in a newborn with seizures and their relationships with multiscale entropy indexes.

Regarding the second analysis related to BHIs by CCM approach, our findings suggest that CNS and ANS are strictly related to ictal events in newborns. This result could help to better understand seizure events and support neonatal seizure detection methods [Olmi et al., 2021]. Preliminary results are promising, but further studies are required exploring other approaches and datasets to check the usefulness of BHI in neonatal seizure detection and characterization. In the next Chapter 5, the ECG/HRV-based NSD methods developed considering all these findings, will be presented.

5. HRV-based methods for neonatal seizure detection

Some contents in this chapter are based on the following publications:

- *Olmi, B., Manfredi, C., Frassineti, L., Dani, C., Lori, S., Bertini, G., Cossu, C., Bastianelli, M., Gabbanini, S., Lanatà, A., 2022. Heart Rate Variability Analysis for Seizure Detection in Neonatal Intensive Care Units. Bioengineering, Vol. 9, Issue 4, p. 165. https://doi.org/10.3390/bioengineering9040165*
- *Frassineti, L., Manfredi, C., Olmi, B., Lanatà, A., 2021. A Generalized Linear Model for an ECG-based Neonatal Seizure Detector. In 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp. 471-474, doi: 10.1109/EMBC46164.2021.9630841.*
- *Frassineti, L., Lanatà, A. Mandredi, C., 2021. HRV analysis: a non-invasive approach to discriminate between newborns with and without seizures. In 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2021, pp. 52-55, doi: 10.1109/EMBC46164.2021.9629741.*

In Chapter 2 and 4 the link between Heart Rate Variability (HRV) and neonatal seizures has been addressed, showing the potentialities, limits and pitfalls of HRV-based approaches in this field. In this chapter this subject will be further exploited, presenting three different approaches regarding the use of HRV as a support to the neonatal seizure detection and characterization.

This chapter is organized as follows:

- In Section 1 the use of HRV measures and Machine-Learning models is presented to quantify the seizure risk, as support to the clinical staff for an ease identification of newborns with seizures.
- Section 2 is devoted to introducing the use of Generalized Linear Model (GLM) and HRV features as HRV-based NSD.
- Section 3 reports the methodologies and the results related to the use of Support Vector Machine (SVM) models as HRV-based NSD.
- Section 4 introduces the MATLAB interface developed during this PhD period, where some of the EEG and ECG techniques described from Chapter 2 to 5, are implemented. The interface was designed according to the clinical needs and physicians' feedbacks collected during this PhD period, supporting the clinical staff in neonatal seizure detection and characterization.

5.1 HRV analysis: a non-invasive approach to discriminate between newborns with and without seizures

As stated in [Statello et al., 2021] and discussed in Chapter 2 and 4, changes in the autonomic nervous system could represent a seizure manifestation and thus a possible neonatal seizures detector. Furthermore, new evidence emerged about links between the autonomic nervous system and neonatal seizures [Olmi et al, 2021, Frassineti et al., 2021d], indicating that the HRV analysis could reveal hidden relevant information. Indeed, these findings suggest that the HRV analysis might be used to discriminate between newborns with seizures and seizure-free ones. Therefore, through the analysis

of EEG or ECG recordings, an index of seizure risk could be found defining a binary classification that highlights the newborns at high risk of seizure events. To better understand this concept, a graphical illustration is provided in figure 5.1.

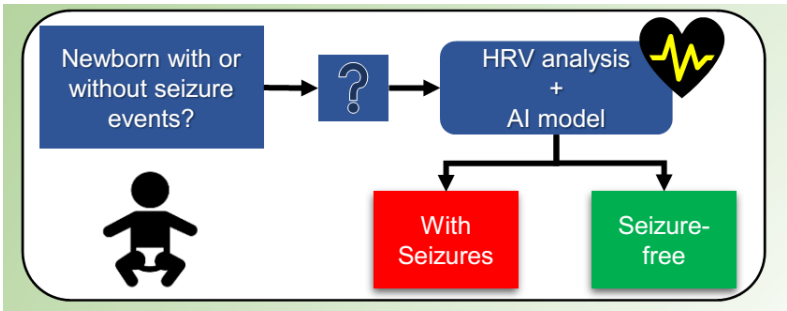


Figure 5.1. Conceptualization of seizure risk index for newborns by HRV analysis and Artificial Intelligence model.

Such systems could be useful in the clinical practice as a pre-screening tool in Neonatal Intensive Care Units (NICUs) to quickly identify newborns that need a deeper neurological investigation by continuous EEG (cEEG) or amplitude EEG (aEEG). To this aim, in this work we investigated if HRV analysis could be effective to detect newborns with seizures.

Preliminary results in seizure detection for adults and children were obtained using time, frequency, and nonlinear HRV features. Here we consider such features as the input dataset of supervised classifiers to recognize newborn with seizures. Proposed methods were trained and validated on the public dataset of neonatal EEG and ECG signals collected in NICU at the Helsinki University Hospital already introduced in Section 2.3 [Stevenson et al., 2019].

5.1.1 Material and Methods

As for the approaches presented in Chapter 4, the methods were defined and tested on 33 patients with consensus seizures and 19 seizure-free subjects from the Helsinki dataset [Stevenson et al., 2019]. To increase the Signal-to-Noise Ratio (SNR) ECGs were pre-processed and filtered with a band-pass FIR filter in the bandwidth 0,05Hz-45Hz. Then the HRV feature set was extracted with the Kubios software version 2.2 [Tarvainen et al., 2014]. Statistical analysis, training and validation of classifiers were implemented under the MATLAB 2019b computing environment. According to [Frassineti et al., 2021d, Statello et al., 2018, Lucchini et al., 2016], for further HRV analysis we defined sliding time windows of 4 minutes of duration, without overlap.

For the HRV analysis, the following feature sets were considered (for a complete description see [Tarvainen et al., 2014, Shaffer and Ginsberg, 2017]):

- Statistical Features: mean of RR intervals (mean_RR); standard deviation of RR intervals (std_RR); mean of HRV (mean_HRV); standard deviation of HRV (std_HRV); root mean square of successive RR interval differences (RMSSD); percentage of successive RR intervals that differ more than 50ms (pNN50); HRV triangular index (HRV_tri_ind); baseline width of the RR interval histogram (TINN).
- Frequency Features: peak, absolute and relative powers of Very Low (VLF), Low (LF) and High Frequencies (HF) using AR models of order 16 for the spectrum estimation

(ARLF_peak) [Tarvainen et al., 2014]; the ratio between LF and HF (ARLF_HF_power), the total power (AR_tot_power) and electrocardiogram derived respiration (EDR).

- Nonlinear Features: Poincarè plot standard deviations (Poincare_SD1 and Poincare_SD2); Approximate and Sample Entropy (ApEn, SampEn with embedding dimension two and tolerance 0.2 [Frassineti et al., 2021d]); Multiscale Entropy (from MSE1 to MSE6, embedding dimension 2 and tolerance 0.2 [Frassineti et al., 2021d]); Detrending short- and long-term Fluctuation Analysis (DFAα1 and DFAα2); Correlation Dimension (CorDimD2).
- Recurrence Plot Analysis Features: Maximum line length (RPALmax); Mean line length (RPALmean); Divergence (RPADIV); Recurrence rate (RPAREC); Determinism (RPADET) and Shannon entropy (RPAShanEn).

According to [Statello et al., 2018], the range of the LF and HF frequency bands was adapted to the neonatal case as follows: LF (0.04-0.3) Hz; HF (0.3-1.3) Hz. According to the time window length and for proper entropy features estimation, multiscale Entropy features were computed up to level 6 (MSE6) [Costa et al., 2005]. Statistical significance of each HRV measure in discriminating between patients with seizures and seizure-free patients was performed as follows:

- Test 1 (T1): Mann Whitney Test (significance level α=0.05) and Permutation Test (number of repetitions 1000, significance level α=0.05) between the medians of the seizure-free patients and the medians of the patients with seizures.
- Test 2 (T2): Mann Whitney Test (significance level α=0.05) and Permutation Test (number of repetitions 1000, significance level α=0.05) between the medians of the seizure-free patients and the medians of the patients with seizures, but considering only the windows containing one or more seizure event (i.e. discarding the interictal time windows).

The workflow for the training and validation of the classifiers was set as follows: only the features relevant to the Permutation Test from T2 were considered, using the features which may discriminate between a window with seizure events and a seizure-free window. Thus, a matrix 52x13 was obtained, where 52 were the patients and 13 the medians of the selected features.

The classifiers were validated through the Leave-One-Subject-Out Validation (LOSO) to avoid overestimation of the performance for the neonatal seizure detection problem [Olmi et al., 2021, Temko et al., 2011a]. Before the validation, the features were normalized (zero mean and unit variances) using the training sets' statistics on the validation sets features. The following machine learning models were trained: linear Support Vector Machine (SVM); Linear Discriminant Analysis (LDA); Random Undersampling Boosting (RUSBoost); k-nearest neighbours (kNN) and Random Forest. Hyperparameter optimization was carried out through the GridSearch method, with the same parameters for each model during the validation procedure. The best model among the set of classifier performance estimates (i.e., accuracy (ACC), $F1_{score}$, Area Under ROC Curve (AUC), Sensitivity (SEN), and specificity (SPE)) was selected, based on the highest average AUC score [Frassineti et al., 2020]. It should be noted that the use of only the significant features from Test T2 might not represent the best subset of features for the classifiers and may lead to overfitting, despite the use of LOSO validation.

Thus, to increase the models' performance, the models were retrained and validated considering both all the features extracted by Kubios and more statistics descriptors besides the median (i.e., mean, standard deviation, maximum, minimum, kurtosis, and skewness) obtaining a matrix of size 52x294. Furthermore, to reduce dimensionality, a feature selection minimum-redundancy-maximum-

relevance algorithm (mRMR) [Ding and Peng, 2005] was implemented, obtaining a final matrix of size 52x29. Afterwards, the same validation procedure was repeated to compare the two approaches. The results and the list of the features considered after the mRMR selection are reported in Section 5.1.2.

5.1.2 Results

Table 5.1 shows the Statistical Tests performed on the 33 patients with seizures compared with the 19 seizure-free subjects: only the features with a significant Permutation Test obtained from T2 are shown. Furthermore, we reported the descriptive statistics mean and standard deviation for all the considered patients. For the patients with seizures, we also reported the values considering only the windows with seizure events. Noteworthy, the significant features are almost the same for both T1 and T2.

Table 5.2 reports the performance of the classifiers, both for the case of features with significant Permutation Test (T2) and that of the mRMR selected features. The Linear SVM with mRMR feature selection showed the highest performance (i.e. 29 predictors). Finally, a list of features selected by the mRMR algorithm is shown in Table 5.3. The threshold for feature selection was empirically given by the highest AUC score.

Table 5.1. Results of Statistical Tests performed on the 33 patients with seizures vs the 19 seizure-free subjects. Only the features with a significant Permutation Test from Test 2 (T2) are reported. The descriptive statistics mean (μ) and standard deviation (σ) are shown. Moreover, for the patients with seizures the statistics of the seizure windows are shown. [Frassineti et al., 2021b].

	Mann Whitney (p-value)		**PermTest (p-value)**		**Patients seizure-free (μ±σ)**	**Patients with consensus seizures (μ±σ)**	
Name Feat	**T1**	**T2**	**T1**	**T2**	**All the windows**	**Windows with seizure**	**All the windows**
std_RR (ms)	0.0109	0.0333	0.0110	0.0280	24 ± 15	15 ± 14	13 ± 12
std_HRV (1/min)	0.0092	0.0333	0.0060	0.0490	7.28 ± 4.29	4.60 ± 4.13	4.01 ± 3.68
HRV_tri_ind	0.0010	0.0046	0.0010	0.0090	6.23 ± 2.87	3.94 ± 2.72	3.59 ± 2.75
TINN (ms)	0.0065	0.0255	0.0050	0.0280	130 ± 70	80 ± 70	70 ± 60
ARLF_power_prc (%)	0.0175	0.0046	0.0190	0.0040	31.70 ± 14.48	20.47 ± 10.92	22.80 ± 10.84
Poincare_SD2 (ms)	0.0087	0.0289	0.0050	0.0310	32 ± 20	20 ± 18	17 ± 16
MSE2	N.S.	0.0383	N.S.	0.0360	0.97 ± 0.27	0.79 ± 0.29	N.S.
MSE3	0.0318	0.0016	0.0390	0.0020	1.06 ± 0.28	0.80 ± 0.27	0.89 ± 0.25
MSE4	0.0166	0.0015	0.0160	0.0020	1.15 ± 0.27	0.88 ± 0.29	0.96 ± 0.26
MSE5	0.0226	0.0062	0.0290	0.0050	1.21 ± 0.28	0.96 ± 0.31	1.03 ± 0.26
MSE6	0.0481	0.0062	0.0579	0.0110	1.26 ± 0.28	1.02 ± 0.32	1.10 ± 0.30
CorDimD2	0.0045	0.0205	0.0220	0.0150	0.58 ± 0.60	0.23 ±0.41	0.23 ± 0.55
RPALmean (beats)	N.S.	0.0098	N.S.	0.0320	20.24 ± 9.21	28.78 ± 16.43	N.S.

Table 5.2. Performance of Leave-One-Subject-Out Validation on the 52 patients: 33 with consensus seizures and 19 seizure-free subjects. On the left: 13 features with significant Permutation Test from T2. On the right: 29 features. [Frassineti et al., 2021b].

	Performances using the 13 significant features to Permutation Test (T2)					**Performances considering all the Kubios features, with more statistics descriptors and MRMR selection (29 features)**				
MODEL	**ACC**	**$F1_{score}$**	**AUC**	**SEN**	**SPE**	**ACC**	**$F1_{score}$**	**AUC**	**SEN**	**SPE**
Linear SVM	65.38%	68.97%	69.86%	60.61%	73.68%	86.54%	89.23%	87.66%	87.88%	84.21%
LDA	63.46%	70.77%	67.30%	69.70%	52.63%	76.92%	81.82%	74.01%	81.82%	68.42%
RUSBoost	65.38%	73.53%	27.27%	75.76%	47.37%	67.31%	72.13%	70.49%	66.67%	68.42%
Random Forest	69.23%	77.14%	65.07%	81.82%	47.37%	63.46%	70.77%	65.07%	69.70%	52.63%
kNN	75.00%	80.60%	60.85%	81.82%	63.16%	80.77%	84.85%	71.93%	84.85%	73.68%

Table 5.3. List of the features selected by the mRMR algorithm. [Frassineti et al., 2021b].

Statistical Descriptor	**Feature Name**
Mean	std_HRV, HRV_tri_ind, ARLF_power
Standard Deviation	HRV_tri_ind, CorDimD2
Median	RMSSD, pNN50, ARVLF_peak, ARLF_power
Max	std_RR, ARVLF_peak, ARVLF_power_prc, AR_tot_power, RPALmean
Min	std_HRV, HRV_tri_ind, ARVLF_peak, ARHF_peak, ARVLF_power, ARLF_power, ARLF_power_prc, ARHF_power, MSE6, CorDimD2
Kurtosis	RMSSD, MSE5
Skewness	ARLF_peak, ARLF_power, ARLF_HF_power

5.1.3 Discussion

Our findings suggest that HRV analysis may successfully catch differences between newborns with seizures and seizure-free newborns in the NICUs. In particular the Linear SVM performance in Table 5.2 suggests that this model is suited to this task, reaching the highest score across the tested models. It is worth noting that the AUC value (about 87%) was obtained using all HRV measures with different statistical descriptors. Moreover, feature relationship analysis through mRMR improved the classification performance with respect to the Permutation Test only. Considering the multiscale entropy features (Table 5.1), lower complexity was found for seizure windows with respect to seizure-free windows. This finding confirms the results already highlighted when entropy indexes were applied on EEG signals during seizure events [Frassineti et al., 2019]. Noteworthy, differences between patients with seizures and seizure-free newborns were evident from the second scale, especially between MSE3 and MSE5, confirming the previous results shown in Chapter 4.

On the contrary, no difference was found without the multiscale analysis (i.e. ApEn, SampEn and MSE1). Furthermore, several analogies were found with [Statello et al., 2018] in the frequency domain features, although the datasets are slightly different (in [Statello et al., 2018] also pre-term newborns were considered). Analogies were found for the total power that was lower for the patients with seizures: mean values of AR_tot_power for the seizure-free subjects were 798 ms^2 while for patients with seizures they were 314 ms^2 (T1 Mann Whitney p-value 0.01). For the HF: the mean

values of ARHF_power for the seizure-free subjects were 73 ms^2 and for patients with seizures they were 43 ms^2 (T1 Mann Whitney p-value 0.03). About the feature selection: although mRMR already gave a consistent improvement of the performance, it did not provide information about the subset relevance. Thus, other methods such as Uncorrelated LDA, Genetic Algorithms; LASSO regression could be used to evaluate different feature subsets and their relevance.

In conclusion, the present study shows the feasibility of HRV analysis as a possible screening tool between patients with and without seizures. Taking into account the lower cost, lower invasiveness, and easier usage of ECG sensors with respect to EEG ones, our findings suggest a possible integration of this approach in NICUs to allow an early detection of newborns at risk of seizures.

5.2 A Generalized Linear Model for an HRV-based Neonatal Seizure Detector

In section 5.1 the index of seizure risk has been introduced, that provides a first screening for clinical staff to detect a newborn with seizures. In this section the development of a HRV-based NSD by Generalized Linear Model (GLM) is presented. Differently from the methods presented in section 5.1, this approach can highlight the time occurrence of the seizure events in the ECG recordings.

As already discussed in Chapter 2, several EEG-based Neonatal Seizure Detectors (NSDs) were proposed in the literature [Olmi et al., 2021]. Moreover, results suggest that Artificial Intelligence techniques could provide a valid support to the clinical staff in the next future [Malak et al., 2019]. Recently, the possibility to develop NSDs without the use of EEG was evaluated [Olmi et al., 2021]. In particular, ECG-based NSDs were proposed in the literature [Olmi et al., 2021]. The idea behind ECG-based NSDs is the use of less invasive, simpler and easily available technologies rather than EEG-based NSDs. Indeed, neonatal seizures may induce direct and indirect alterations to the autonomic nervous systems, thus they could be better detected by ECG-based NSDs [Statello et al., 2021]. Unfortunately, the performance of these detectors is still too low to represent a valid alternative to EEG for newborns [Statello et al., 2021, Olmi et al., 2021].

In fact, most of the computer-based systems proposed in the literature are developed to automatically detect seizure activity in adults and children. De Coomant et al. proposed a patient-independent algorithm for online epileptic seizure detection based on the analysis of single-channel ECG data from eight patients (age: 29–51 years) with temporal lobe epilepsy [De Cooman et al, 2017]. They used Linear Support Vector Machine and Linear Discriminant Analysis (LDA) classifiers. The best performance was found for the LDA-based system, giving a Sensitivity (SEN) = 80% and Specificity (SPE) = 87%. Behbahani et al. proposed Multilayers Perceptron (MLP) neural networks with different numbers of hidden layers to detect seizures, performing the HRV analysis on ECG data from 15 patients (9 patients with complex partial seizures, 6 with secondarily generalized seizures, mean age 42.2 years) [Behbahani et al., 2014]. The MLP trained with the Levenberg–Marquardt algorithm gave SEN = 83.33%, SPE = 86.11 and Accuracy (ACC) = 84.72% for the patients with complex partial seizures, and for the patients with secondarily generalized seizures, SEN = 86.66%, SPE = 90% and ACC = 88.33%.

Jeppesen et al. validated an HRV-based seizure detection algorithm analyzing ECG signals recorded using a wearable device [Jeppesen et al., 2019]. They performed an offline and patient-specific analysis based on recordings from 19 patients (age: 4–62 years). The overall SEN was 56.3%. They observed that the algorithm performed better on patients with marked autonomic changes, giving an SEN = 87%.

Although these methods show an appealing performance, they cannot be directly applied to newborns due to their quite different electrophysiological activity. However, some attempts to support seizure diagnosis through the ECG/HRV analysis have also been made for newborns.

Greene et al. presented an ECG-based system using a Linear Discriminant (LD) classifier [Greene et al. 2007b]. They considered a dataset of ECG recordings from seven full-term newborns suffering from hypoxic ischemic encephalopathy (HIE) in the NICU of the Unified Maternity Hospitals in Cork, Ireland, and the Kings' College Hospital, London. The HRV analysis was performed extracting properties of the R-R intervals in time, frequency and information theory domains. They developed a patient-specific and a patient-independent system. The first one provided Accuracy (ACC) = 66.04%, Sensitivity (SEN) = 75.52% and Specificity (SPE) = 57.70%, while the patient-independent system gave ACC = 61.80%, SEN = 78% and SPE = 51.75%. Doyle et al. investigated the usefulness of the HRV analysis to detect neonatal seizures introducing a Support Vector Machine (SVM)-based system [Doyle et al., 2010]. They considered the ECG recordings of 14 full-term newborns admitted in the NICUs of the Unified Maternity Hospitals in Cork, Ireland. Concerning the HRV analysis, only a feature set defined in time and frequency domains was considered. The system results were: mean Area Under the ROC Curve (AUC) = 60% and mean SEN = 60%

However, progresses in nonlinear HRV analysis in newborns and recent findings on neonatal seizures [Lucchini et al., 2016, Frassineti et al., 2021d] opened up new perspectives for the improvement of these methodologies. Therefore, further studies should be carried out to investigate the role of HRV in neonatal seizure detection and to check other methods that might be better suited for NSD than the already existing ones. To this aim, Generalized Linear Models (GLMs) could be a promising approach. GLMs were proposed to detect seizures in rats using features extracted from Electroencephalographic (EEG) signals [Fumeaux et al., 2020] or to classify normal and pre-ictal EEG signals [Redelico et al., 2017]. Indeed, these models are widely used in several biomedical fields thanks to their simplicity in describing the relationships between measured variables and outcomes [Redelico et al., 2017].

The present work evaluates if GLM models may also be valuable as HRV-based NSDs. To this end, this study provides the first proof of concept about the development and use of GLMs to detect and characterize neonatal seizures through HRV measures. If successful, it might enable the application of such seizure detectors for newborns or infants also in-home monitoring environments or when EEG-based techniques are not readily available.

5.2.1 Material and Methods

As for the approaches presented in Chapter 4 and section 5.1, the methods presented here were developed and tested on 33 patients with consensus seizures and 19 seizure-free subjects from the Helsinki dataset [Stevenson et al., 2019]. Moreover, the same pre-processing as in Section 5.1 for ECG was applied. The same HRV feature set proposed in Section 5.1.1 was extracted with the Kubios software version 2.2 [Tarvainen et al., 2014]. Statistical analysis and the GLM model validation were implemented under the MATLAB 2020b computing environment. For the HRV analysis, for each recording non-overlapping sliding time windows lasting 4 minutes were defined [Statello et al., 2018, Lucchini et al., 2016]. Artifacts were removed using a first-order detrending step and a "medium correction". For more details, see [Tarvainen et al., 2014]. Then a Mann-Whitney Test (significance level α=0.05) between the medians of the windows of the seizure-free patients and those with one or more seizure event was performed. This test aimed at assessing the statistical significance of HRV

measures to discriminate windows with seizures events from seizure-free windows. The relevant features found after Mann-Whitney tests are shown in Table 5.4. Table 5.4 shows only the subset of significant HRV features from those already introduced in section 5.1.1. For a detailed description of them please see [Shaffer and Ginsberg, 2017, Frassineti et al., 2021b, Frassineti et al., 2021c].

To implement the GLM model for the NSD, a stepwise regression procedure was performed. Starting from a model with only the intercept term and considering the subset of significant features found with the Mann-Whitney test, we used a forward and backward stepwise regression to determine the final model. The criterion used to add or remove terms was the Deviance Criterion [Myers and Montgomery, 1997]. Moreover, the GLM model was trained using the Binomial Distribution for the response variable and with a Logit link function [Myers and Montgomery, 1997].

Table 5.4. Significant HRV features after the Mann-Whitney test regarding differences between windows with seizures and seizure-free windows (level of significance 0.05). The overall list of HRV features is presented in Section 5.1.1. [Frassineti et al., 2021c].

Feature Name	p-value	Feature Name	p-value
std_RR	0.03	MSE3	0.001
std_HRV	0.03	MSE4	0.001
RMSSD	0.01	MSE5	0.006
HRV_tri_ind	0.004	MSE6	0.006
TINN	0.02	CorDimD2	0.02
AR_LF_power	0.01	RPA Lmean	0.009
AR_LF_power_prc	0.004	RPA REC	0.04
AR_HF_power	0.02	RPA ShanEn	0.02
MSE2	0.03		

The GLM model was built using all the considered time-windows, including the interictal time windows that were not used for the statistical test. In total, 1067 windows from the 52 patients were used, 284 of which with seizure events. As in the classical binary seizure classification problem, seizure-free epochs were labelled with “0” and epochs with seizure events with “1”.

Before the stepwise procedure, the features were normalized by rescaling the data range in the interval [0,1], where 0 is the lowest value of the features across all windows and 1 is the highest value. Furthermore, missing values were replaced with the overall median values of the features. From the model, the concatenated Area Under the ROC curve was computed (AUC_{cc}) [Frassineti et al., 2020] to assess the model's performances in detecting windows with seizure events. Furthermore, a Leave-One-Subject-Out Validation (LOSO) was defined, iteratively removing each patient and retraining the GLM model using the same formula obtained by the stepwise regression procedure. LOSO validation was applied to avoid overestimation of the neonatal seizure detection task [Temko et al., 2011a]. The following patient-independent performances were defined: Accuracy (ACC), Sensitivity (SEN), Specificity (SPE). Performances were obtained after the selection of the threshold parameter for the response variables. Results are presented in Section 5.2.2.

5.2.2 Results

The final model's formula obtained from the stepwise procedure is shown in a concise symbolic form in Equation 5.1 where only the interaction terms for RMSSD, MSE3 and MSE5 are displayed:

$$labels \sim 1 + ARLF_{power} + RMSSD \times MSE3 + RMSSD \times MSE5 \quad (5.1)$$

Table 5.5 shows the statistical results related to the GLM model. Both the nonlinear entropy features MSE3 and MSE5, and their interaction with RMSSD show a statistically significant p-value (significance level α=0.05). Besides, we tested if the model significantly differs from a constant model using a Deviance Test. Constant model was defined as a model that considers only the intercept term without any independent variables, i.e. the HRV features considered. A χ^2 statistic vs constant model: 96.6 and p-value 1.3e-18 was obtained.

In Figure 5.2, the partial dependence plots of the predicted labels as a function of the predictor variables involved in the interaction terms (RMSSD, MSE3 and MSE5) are shown. Partial Dependence is defined as the relationships between predictor's variables and predicted labels [Friedman, 2001]. Values in the colormap close to 1 represent windows with seizures, otherwise values close to 0 windows seizure-free. From Figure 5.2a and b it is possible to denote that there are two regions in the Partial Dependence Plots where the values are close to 1, corresponding to the extreme values of RMSSD, MSE3 and MSE5.

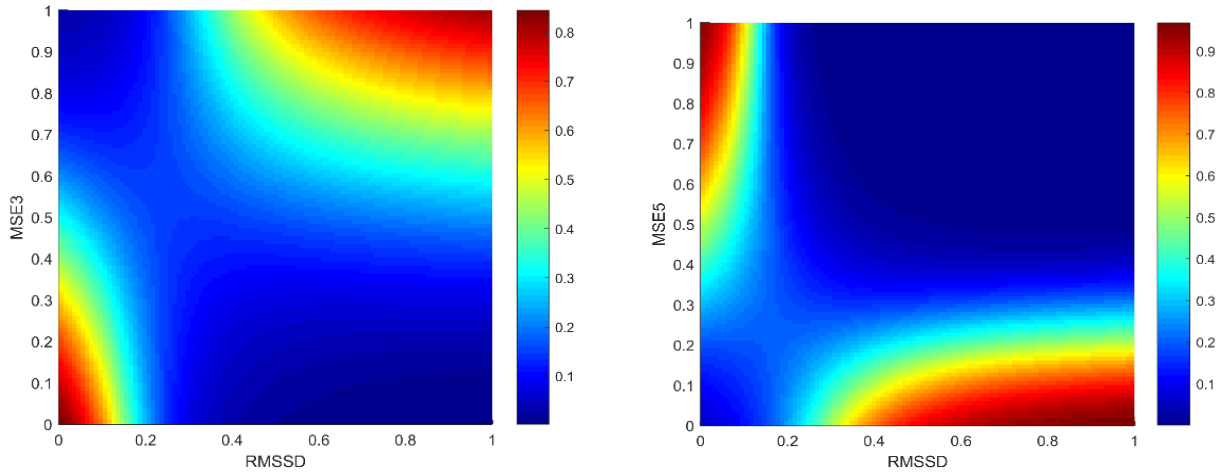
Table 5.5 Estimated coefficients of the proposed GLM model. The model was built using 1067 windows from 33 patients with seizure events and 19 seizure-free subjects. [Frassineti et al., 2021c].

	Estimate	**SE**	**tStat**	**p-value**
Intercept	-0.33107	0.25393	-1.3038	0.19
RMSSD	-0.63561	1.6911	-0.37585	0.70
ARLF_power	-162.77	87.707	-1.8558	0.06
MSE3	-6.116	1.1778	-5.1928	2.07e-07
MSE5	6.7515	1.4574	4.6324	3.61e-06
RMSSD:MSE3	25.728	6.5232	3.9441	8.01e-05
RMSSD:MSE5	-42.402	9.9384	-4.2664	1.98e-05

In Table 5.6, the results of the GLM model as an ECG-based NSD are shown. For patient-independent metrics ACC, SEN and SPE, the threshold (TH) used to obtain these performances and their mean values with standard deviations obtained after the LOSO validation on the 52 patients are reported. The chosen TH value allowed a good compromise between SEN and SPE. The ROC curve related to the parameter AUC_{cc} is shown in Figure 5.3.

Table 5.6 Performances of the proposed GLM model. [Frassineti et al., 2021c].

Method	**AUC_{cc} (%)**	**TH**	**ACC (%)**	**SEN (%)**	**SPE (%)**
GLM (LOSO)	69.69	0.35	68±27	43±37	77±28



(a) (b)

Figure 5.2 (a) Partial Dependence Plot of normalized RMSSD and MSE3 as a function of the predicted labels. (b) Partial Dependence Plot of normalized RMSSD and MSE5 as a function of the predicted labels. [Frassineti et al. 2021c].

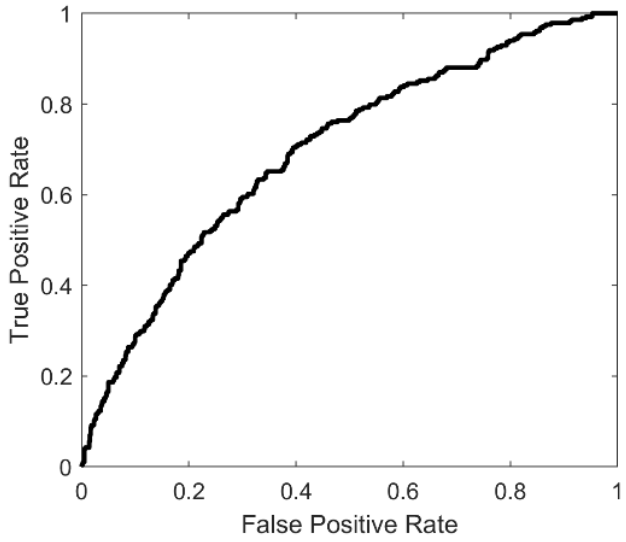


Figure 5.3 ROC curve of the AUC_{cc} value reported in Table 5.6. [Frassineti et al., 2021c].

5.2.3 Discussion

This work aims at evaluating if HRV analysis and GLM models could allow the development of an ECG-based NSD. The results reported in Table 5.5 show that the Multiscale Entropy contributes to characterize the seizure events with the model obtained. This possibility was not previously exploited in NSDs [Olmi et al., 2021]. The obtained performances confirm that EEG-based NSDs are still better than the ECG ones [Olmi et al., 2021]. However, the AUC_{cc} obtained by the GLM model can be considered a relevant improvement for the development of ECG-based detectors [Greene et al., 2007b, Doyle et al., 2010]. Furthermore, as reported in Section 5.2.1 and as in [Statello et al, 2018], significant differences were found for the HF features, where the Mann-Whitney tests gave p-values lower than 0.05. Significant differences were also found for the LF features. Considering the interactions shown in Figure 5.2, it is interesting to highlight the information provided by RMSSD values: extreme values of this measure (very close to 0 and 1) reflect an abnormal parasympathetic

activity inside the windows with seizure events. At the same time, at different scales, the MSE analysis (Figure 5.2) captures useful information about abnormal heart rate dynamics related to seizure events, e.g. reduced variability or transient decelerations [Frassineti et al., 2020d, Costa et al., 2005]. Table 5.5 shows that the multiscale approach (from MSE2 to MSE6) allows finding statistical differences between the ictal and the seizure-free periods that could not be detected with a single scale approach (by ApEn or MSE1/SampEn). Furthermore, low values of the entropy indexes during ictal events are similar to the EEG case [Frassineti et al., 2019].

The results confirm that neonatal seizures may alter the cardio-regulatory system and an ECG-based NSD may detect these changes. It was already demonstrated [Bersani et al, 2020] that HRV analysis can provide a reliable marker of brain damages in the case of Hypoxic-Ischaemic Encephalopathy (HIE), the most common aetiology behind neonatal seizures [Pressler et al, 2021]. However, this finding cannot be extended to all the newborns and seizures events considered. As shown in Table 5.6, the high standard deviations obtained in LOSO validation mean that these alterations were not present for some patient, or the used HRV features cannot detect them. This is probably due to the possible different kind of seizures' aetiologies [Pressler et al., 2021]. Another possibility is that HRV analysis may be unspecific for the neonatal seizure detection problem. For example, motor activity during ictal events could lead to changes in heart rate and its variability, although often neonatal seizures are electrographic-only [Pressler et al., 2021]. Furthermore, it is well known that hypothermia or pharmacological treatment may alter both EEG and HRV analysis. However, for the HD dataset this information was not provided thus it was not possible to evaluate such factor. Moreover, the CD dataset was collected retrospectively, and for some patients it was not known when the treatment was applied (before, during or after the EEG/ECG analysis). Therefore, the analysis of possible alterations due to other factors than seizures in the HRV signals or the GLM predictions could not be included in the presented work. Therefore, further analysis is needed to confirm the use of only HRV measures in NSDs. This could be achieved setting up a specific dataset that collects and monitors the effects of treatments on the subjects. Also, taking into account different aetiologies for neonatal seizures rather than just HIE could provide useful information. Moreover, our NSD is based on windows lasting 4 minutes that cannot detect the exact temporal occurrence of seizures as their average duration is often lower than the window used [Stevenson et al., 2019], whose time duration is about 80-100 seconds on average for this dataset. In other words, the proposed NSD can detect the windows containing one or more seizure events but cannot establish their exact onset and offset. This is a trade-off due to the limitations of the HRV feature extraction method that makes use of short time windows [Shaffer and Ginsberg, 2017, Lucchini et al., 2016]. However, other ECG-based NSDs could be developed in the next future based on short-time windows (i.e. less than 30-60 seconds). The proposed approach represents a valid support for the clinical decision process to detect newborns' ictal periods, capable to highlight only the periods with seizures and thus reducing the number of recording hours to be inspected by the physician. In conclusion, considering the low invasiveness, low cost, and easier usability of ECG sensors with respect to EEG ones, our results suggest a possible integration of these systems in NICUs or any situation where EEG technologies are not easily and timely available.

5.3 Heart Rate Variability Analysis and Support Vector Machine for Seizure Detection in Neonatal Intensive Care Units

In this section the application of HRV and AI models as Neonatal Seizure Detectors is presented. Since Support Vector Machine (SVM) models were widely used in EEG-based NSD [Olmi et al.,

2021], in this work it has been investigated if such supervised classifiers could be used also as ECG-based NSD. In this work the CD dataset was used for all the methods implemented. This dataset was already described in section 2.3 but its features are reported here for ease of reading. From the Careggi dataset, 51 full-term newborn (22 with seizure events) were considered. More specifically, 10 out of the 22 pathological newborns showed electrographic-only seizures (EGP) characterized by abnormal changes in the EEG signal and poor clinical signs such as ocular, oral/buccal/lingual and progression movements [Volpe, 1989, Mizrahi and Kellaway, 1987]. The remaining 12 subjects exhibited electroclinical seizures (ECP), characterized by clinical signs coupled with EEG changes [Pressler et al., 2021]. None of the considered newborns have heart disease that could affect the study; thus, the proposed analysis was performed on the whole dataset. The mean length of recordings per patient was 53 minutes, the overall duration of the dataset was about 45 hours. The mean seizure duration per patient was 00:09:39 h. The sampling frequency of ECG was 128 Hz. The ECG signals were pre-processed using a high-pass filter with a time constant of 0.1 s and a 50 Hz Notch filter.

5.3.1 Methods

This section describes the methods implemented to develop an SVM-based system. SVM makes use of both the Gaussian and the Linear kernel. The ECG signals were segmented into non-overlapping time windows, called "epochs" [Olmi et al., 2021]. Both 60 s epochs [Doyle et al., 2010] and 180 s epochs [Statello et al., 2018] were considered. An experienced neurologist labelled the seizure events indicating both the beginning and the ending time instant of each seizure. Since an event labelled by the clinician did not precisely overlap with one or more epochs, we labelled an epoch as a "seizure epoch" if at least one sample of the signal falls inside a time interval previously classified by the experienced neurologist as a seizure event. One of the most crucial aspects of the ECG analysis was the identification of QRS complexes for the segmentation into single beats. This ECG pre-processing enabled HRV estimation and thus found out the possible seizure's effects on the ANS [Statello et al., 2021]. In particular, the localization of the R peaks was performed by implementing the Pan–Tompkins' algorithm [Pan and Tompkins, 1985], as it was previously used for R peaks detection from neonatal ECGs [Frassineti et al., 2021b, Frassineti et al., 2021c]. Given the time instants at which R peaks were detected, we computed the time distance between each pair of consecutive R peaks obtaining the RR time series. In order to develop a fully-automatic HRV-based NSD system, no correction procedures based on visual inspection were performed to remove false detections and ectopic beats. The HRV analysis was performed for each epoch, extracting a set of features defined in time, frequency and information theory domains. More specifically, we extracted 18 features for both 60 and 180 s epochs. These features were selected among the most widely used in the literature for HRV analysis [Shaffer and Ginsberg, 2017, Lucchini et al., 2016]. To increase the system performance, we introduced additional features concerning the multiscale entropy, as they could provide additional useful information to characterize alterations in cardiovascular activity during neonatal seizures [Frassineti et al., 2020d]. More specifically, we considered the Multiscale Sample Entropy (MSE) [Costa et al., 2005] and the Multiscale Distribution Entropy (MDE) [Lee et al., 2018], which were computed implementing the coarse-grained procedure [Costa et al., 2005]. The MSE was implemented defining the embedding dimension equal to 2 and the tolerance value equal to 0.2 [Frassineti et al., 2020d]. The MDE was implemented defining the embedding dimension equal to 2 and the number of bins equal to 512. Moreover, considering the average newborn heart rate at rest (100–200 bpm [Doyle et al., 2010]), we computed these features up to scale 4 for the 180 s epochs and considered only the scale 1 for the 60 s epochs. This allowed achieving at least 10^2 points on each scale. This choice was made to avoid an inaccurate estimation of entropy parameters due to a coarse-

grained scale at higher scales where the number of points could be too low [Frassineti et al., 2020d]. All the considered features and a short description of each of them are reported in Table 5.7.

Table 5.7. Features extracted for HRV analysis. Domain, unit of measure and a short description are provided for each feature. [Olmi et al., 2022a].

	Feature	Unit of Measure	Short Description
Time domain	SDSD	(ms)	Standard deviation of successive R-R interval differences
	SDNN	(ms)	Standard deviation of R-R intervals
	RMSDD	(ms)	Root mean square of successive differences
	pNN50	(%)	Probability of R-R intervals > 50 ms e < −50 ms
	TRI	-	Area of the histogram of R-R intervals divided by its maximum height
	TINN	(ms)	Width of the R-R intervals histogram evaluated trough triangular interpolation
	CD	-	Correlation dimension
	SD2	(ms)	Standard deviation of Poincarè plot along the line-of-identity
	SD1SD2ratio	-	Ratio of standard deviation of Poincarè plot perpendicular to the line-of-identity to standard deviation of Poincarè plot along the line-of-identity
	HR	(beats/min)	Average heart rate
Frequency domain	VLF	(ms^2)	Spectral density (computed through FFT) of the linear interpolated R-R tachogram up to 0.04 Hz (very low frequency) [25]
	LF	(ms^2)	Spectral density (computed through FFT) of the linear interpolated R-R tachogram between 0.04 and 0.3 Hz (low frequency) [25]
	HF	(ms^2)	Spectral density (computed through FFT) of the linear interpolated R-R tachogram between 0.3 and 1.3 Hz (high frequency) [25]
	LFHFratio	-	Ratio between spectral density of low frequency parts and high frequency parts
	TP	(ms^2)	Total spectral density
	pLF	(%)	Percentage of spectral density of low frequency parts to total spectral density minus the spectral density of very low frequency parts
	pHF	(%)	Percentage of spectral density of high frequency parts to total spectral density minus the spectral density of very low frequency parts
Information theory domain	ApEn	-	Approximate Entropy
	Multiscale DistEn Scale (1–4)	-	Multiscale Distribution Entropy from scale 1 to scale 4 for the 180 s epochs; at scale 1 for the 60 s epochs
	Multiscale SampEn Scale (1–4)	-	Multiscale Sample Entropy from scale 1 to scale 4 for the 180 s epochs; at scale 1 for the 60 s epochs
Total	20 (60 s epochs)/26 (180 s epochs)		

The feature sets extracted from each epoch were normalized (zero mean and unit variance). To manage the presence of missing feature values in the dataset, an imputation process was performed. It replaces missing values with the average values of the features set. Moreover, a feature selection algorithm was applied to identify the most informative subset of features. The feature selection process allows reducing the number of features that did not add significant information, making the classes separation difficult. The Minimal-Redundancy-Maximal-Relevance (mRMR) algorithm was implemented [Ding and Peng, 2005]. The mRMR algorithm maximized the mutual information between the features and the target class. It ensured that the mutual information between the new features and the already chosen ones was minimal. The features corresponding to the minimum redundancy and the maximum relevance were selected. The optimal features set selection step was performed by comparing the classification performances of a Support Vector Machine (SVM) classifier fed by the full features set and then the subsets of 2 to 20 features chosen according to the order of ranking determined with mRMR.

The Linear SVM is a supervised learning technique that performs classification, finding the hyperplane that maximizes the margin between the two considered classes, thus separating data into two non-overlapping classes [Tong and Koller, 2002]. When data are not perfectly separable, SVM searches for the hyperplane that maximizes the margin and minimizes the misclassifications by introducing a regularization penalty term called λ. In general, the problem of maximizing the margin leads to minimizing the norm of the vector perpendicular to the hyperplane. This optimization problem can be solved using different routines called Solvers. When data are not linearly separable, kernel functions are applied to map the samples into a high-dimensional feature space in which linear

classification is possible. In that case, the Gaussian kernel approach was used. The main hyperparameters of the Linear and the Gaussian SVMs with a short description are reported in Table 5.8. Seizure detection is an inevitably unbalanced problem because a short duration usually characterizes critical events as compared to non-critical activity [Olmi et al., 2021]. In the CD, the total duration of pathological patients' recordings was 23:37:58 h, and the total duration of their seizure activity was only 03:32:11 h. Moreover, we also considered control patients, thus making the dataset more unbalanced. The unbalanced data distribution can affect the detection and classification performance leading to biased classifier results toward the majority class to which the non-seizure data belong. To deal with this problem, data in the non-seizure and the seizure class were managed asymmetrically by introducing the Costs C_1 and C_2, respectively, thus assigning different weights to the elements of the classes during the training step [Awad and Khanna, 2015].

Table 5.8. Main hyperparameters of the Linear and Gaussian SVM. [Olmi et al., 2022a].

Linear SVM	
Hyperparameters	**Short Description**
λ	Regularization penalty term introduced to search for the hyperplane that maximizes the margin and minimizes the misclassifications.
Costs	Misclassification costs introduced to mitigate the class imbalance that occurs when one class has a smaller number of examples with respect to the other.
Gaussian SVM	
Box Constraints	Regularization term that controls the number of misclassifications.
Kernel Scale	Scaling parameter for the input data preventing some features that have a wider range than others from becoming dominant in the kernel calculation.
Costs	Same definition of Costs for Linear SVM

To find the optimal hyperparameters values for each classifier, we performed the Grid Search optimization and the Leave One-Subject Out (LOSO) cross-validation. The Grid Search operation implemented an exhaustive search through a manually specified subset of the hyperparameter space of the learning algorithm. The LOSO method provided an almost unbiased estimation of the true generalization error. It is an iterative method: at each iteration, the training set is defined by excluding one patient, and the test set is composed of the data from that excluded patient. This process is repeated until each patient has been considered as a test subject. Thus, LOSO performs a good evaluation of the system's ability to generalize the classification: once trained on all the available data, it achieves performances similar to those obtained by the system with an unknown dataset [Olmi et al., 2021, Temko et al., 2011a]. LOSO is very useful for small datasets as no subsampling of the original dataset is performed, thus reducing the risk of overfitting. The Grid Search operation was implemented through a model for every combination of specified hyperparameters. Then, the following metrics were calculated:

- Epoch-based metrics:
  - Area Under the ROC Curve (AUC)
  - Sensitivity (SPE), Specificity (SPE)
  - $F1_{score}$ (F1)
- Event-based metrics:
  - Good Detection Rate (GDR)
  - False Detection per hour (FDH)

- False Discovery Rate (FDR)
- Time Delay

A detailed description of such metrics is reported in sections 1.3.3 and 2.2.1. More specifically, Linear and Gaussian SVMs were trained using the full set of HRV features and the subsets selected through mRMR. The final models were selected as the ones with the best average AUC values for pathological patients. Performance and hyperparameters of these models, both using 60s and 180s epochs, will be presented in section 5.3.2.

5.3.2 Results

In table 5.9 the LOSO SVMs' performance of the models with the best AUC obtained in the experiment based on ECG epochs of 60s are reported, while in Table 5.10 the LOSO SVMs' performance of the experiment based on the ECG epochs of 180s are shown.

Table 5.9. Models with the best average AUCs obtained in the experiment based on the segmentation of the ECG signal into 60 s implementing the LOSO cross-validation. [Olmi et al., 2022a].

Model	N° Features	Hyperparameters	AUC (%)	SEN (%)	SPE (%)	GDR (%)	FDH (h^{-1})	FDR (%)	F1 (%)	Time Delay (s)
			(Mean ± Standard Error)							
Linear SVM	Full feature set (20)	$\lambda = 10^{-5}$ Solver: dual $C_1 = 1; C_2 = 2$	52 ± 4	24 ± 7	89 ± 3	27 ± 8	2 ±1	4 ± 1	12 ± 4	56 ± 3.5
	Features selected through mRMR (20)	$\lambda = 10^{-8}$ Solver: dual $C_1 = 1; C_2 = 7$	54 ± 3	26 ± 8	87 ± 3	36 ± 8	3 ± 0.4	5 ± 1	15 ± 4	116 ± 10
Gaussian SVM	Full feature set (20)	Box Constraint: 1 Kernel Scale: 5 $C_1 = 1; C_2 = 7$	52 ± 3	29 ± 8	84 ± 4	34 ± 9	4 ± 1	6 ± 1	16 ± 5	42 ± 0.6
	Features selected with mRMR (5)	Box Constraint: 0.5 Kernel Scale: 1 $C_1 = 1; C_2 = 7$	54 ± 3	24 ± 8	85 ± 2	27 ± 9	3 ± 1	6 ± 1	16 ± 5	55 ± 3

Table 5.10. Models with the best AUCs obtained in the experiment based on the segmentation of the ECG signal into 180 s implementing the LOSO cross-validation. [Olmi et al., 2022a].

Model	N° Features	Hyperparameters	AUC (%)	SEN (%)	SPE (%)	GDR (%)	FDH (h^{-1})	FDR (%)	F1 (%)	Time Delay (s)
			(Mean ± Standard Error)							
Linear SVM	Full feature set (26)	$\lambda = 10^{-7}$ Solver: dual $C_1 = 3; C_2 = 1$	56 ± 5	22 ± 7	87 ± 3	31 ± 8	1 ± 0.2	6 ± 1	20 ± 6	141 ± 4
	Features selected with mRMR (2)	$\lambda = 10^{-7}$ Solver: dual $C_1 = 1; C_2 = 40$	58 ± 5	22 ± 9	77 ± 5	25 ± 9	1 ± 0.2	4 ± 1	13 ± 1	138 ± 15
Gaussian SVM	Full feature set (26)	Box Constraint: 0.5 Kernel Scale: 25 $C_1 = 1; C_2 = 5$	50 ± 4	51 ± 1	61 ± 5	58 ± 10	2 ± 0.3	10 ± 1	27 ± 6	117 ± 13
	Features selected through mRMR (2)	Box Constraint: 5 Kernel Scale: 0.1 $C_1 = 1; C_2 = 200$	62 ± 5	47 ± 8	67 ± 3	62 ± 9	3 ± 0.3	16 ± 1	29 ± 5	123 ± 3

The results reported in Tables 5.9 and 5.10 suggest that an epoch length of 180 s is appropriate to analyze and detect seizure events in the considered dataset, because with 180s the highest AUC values were obtained. Therefore, we focused on the experiment based on the signal segmented into 180 s epochs. Figure 5.4 shows the 26 features defined for this experiment and described in Table 5.10. It also displays their classification relevance based on the mRMR's predictor importance score.

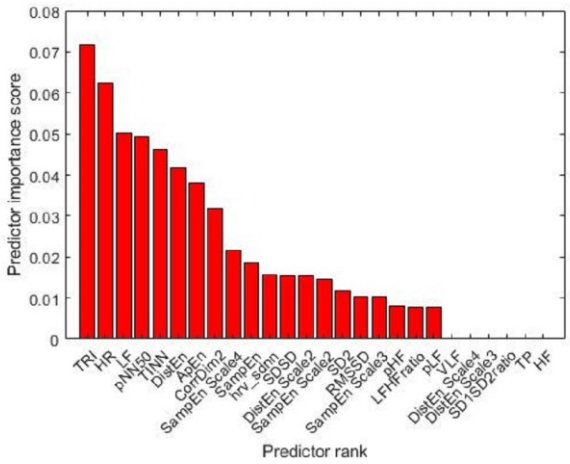


Figure 5.4. The features ranked with the mRMR algorithm for the experiment based. [Olmi et al., 2022a].

Moreover, we evaluated the Sensitivity values for each pathological patient. The following notation is considered: ECP = newborns with electroclinical seizures, EGP = newborns with electrographic seizures. More specifically, the model trained using the full set of features resulted in 10 out of 22 pathological patients characterized by Sensitivity values >0. The model trained with the subset of 2 features selected through the mRMR algorithm, gave worse results: only 7 out of 22 pathological patients were characterized by Sensitivity values >0. Similarly, the Gaussian model trained with the full set of features gave 15 patients with Sensitivity values >0, instead the Gaussian model trained with the subset of two features selected with the mRMR algorithm gave 17 patients with Sensitivity values >0. Considering the overall performance, the Gaussian SVM model trained using a subset of two features seems to provide a good tradeoff between a high AUC value (mean ± standard error: 62 ± 5%) and a large number of patients with Sensitivity >0. We also calculated the concatenated AUC (AUC_{cc}), defined as the Area Under the ROC curve, across all the concatenated recordings [Tapani et al., 2019]. This ROC curve, shown in Figure 5.5, was built linking together all the recordings from control and pathological patients. The AUC_{cc} was equal to 63%.

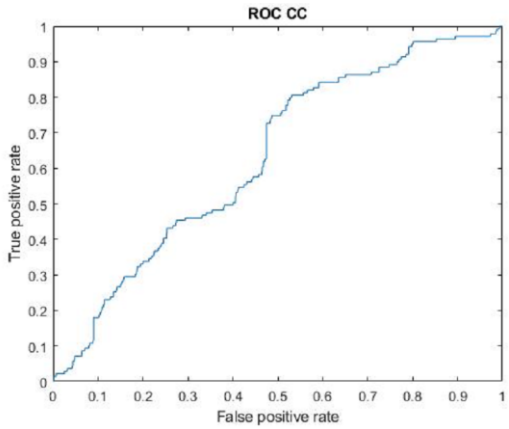


Figure 5.5. The concatenated ROC evaluated on the SVM Gaussian kernel-based system with highest AUC. All the recordings are linked together into a single recording. [Olmi et al., 2022a].

5.3.3 Discussion and Conclusions

This study aimed at developing an SVM-based system to automatically detect neonatal seizures in NICUs by investigating ECG recordings. In order to analyze the effects that seizures might have on

the Autonomic Nervous System (ANS), the HRV analysis was performed based on features defined in time, frequency and information theory domains.

The results reported in Tables 5.9 and 5.10 suggest that, for CD, an epoch length of 180s achieved better ability in detecting and analyzing seizure events than 60 s epochs, although it leads to an increase in the seizure detection delay, which is assessed with the time delay metric. Indeed, this measure is heavily influenced by the time duration of the epochs in which the signal is segmented and the processing time required to run the algorithms. In the experiment based on the signal segmentation into 60 s epochs, the time delay ranged between 42 ± 0.6 and 116 ± 10 s (mean ± standard error). In the experiment based on the signal segmentation into 180 s epochs, the time delay ranged between 117 ± 13 and 141 ± 4 s (mean ± standard error). Thus, although as shown in Table 5.10, it achieved a better overall performance, the system developed segmenting the ECG signals into 180 s epochs does not seem suitable for real-time applications. However, the use of overlapping time windows, that could reduce the system's response time, should be investigated.

As mentioned above, we labelled an epoch as a seizure epoch if at least one sample of the signal falls inside a time interval previously classified by the experienced neurologist as a seizure event. However, finding a single clinical definition of neonatal seizure is still challenging, thus finding an operational definition for the NSD task is also very tricky. Some papers consider a seizure epoch when 50% of the time window contains a seizure [Olmi et al., 2021]. This choice is not feasible for our dataset segmented into 180 s epochs, because about 66% of seizures last less than 90 s, with the consequence of losing some of them. According to the American Clinical Neurophysiology Society (ACNS) [Tsuchida et al., 2013], that defined an electrographic neonatal seizure as "*a sudden, abnormal EEG event, defined by a repetitive and evolving pattern with a minimum 2 μV peak-to-peak voltage and duration of at least 10 s*", we tested our methods re-labeling an epoch as a "seizure epoch" if it contains at least 10 s of seizure previously classified by the experienced neurologist as a seizure event. However, we achieved lower performance than those reported in Tables 5.9 and 5.10. Indeed, ACNS also remarks that the choice of 10 s is conventional and arbitrary; thus, other seizure durations should also be evaluated in NSD experiments. Based on this evidence, the choice of considering at least one signal sample could be cautionary, even if drastic.

The AUC values reported in Table 5.10 show that the Linear and Gaussian SVM models trained using a subset of two features look promising. They gave AUC values equal to 58 ± 5 and 62 ± 5% (mean ± standard error). These models were trained using the Triangular Index (TRI) and the average heart rate measurements, as shown in Figure 6. The TRI is a geometrical measure that estimates the overall HRV [Camm et al., 1996]. Its main advantage is its relative insensitivity to artefacts [Hämmerle et al., 2020]. To the best of our knowledge, no other studies analyzed the relationship between the TRI values and the seizure events in newborns. However, some works developed for adults [Yıldız et al., 2011] and children [Kolsal et al., 2014] have observed that TRI values were lower in epileptic patients than in controls, reflecting reduced parasympathetic and increased sympathetic activities during seizures [Hämmerle et al., 2020]. This research evidence supports our results, suggesting that TRI is a robust and valuable measure for discriminating between seizure and non-seizure epochs.

The average heart rate feature is important as well. Indeed, heart rate is one of the most common autonomic manifestations of seizures in the neonatal period [Pressler et al., 2021]. Several authors reported significant changes in heart rate during adult epileptic seizures [Akyüz et al., 2021]. These changes were also observed in neonatal seizures. Greene et al. considered five recordings from four epileptic neonates, and their study showed a significant increase in heart rate during clinical and

subclinical seizures [Greene et al., 2006]. Statello et al. found that HR values were significantly higher during seizure events than the interictal periods far from the seizure events [Statello et al., 2018].

The Gaussian SVM model trained using the two features (TRI and average HR) should be preferred to the Linear one, although it gave worse values of FDH (mean ± standard error: 3 ± 0.3 h^{-1}) and FDR (mean ± standard error: 16 ± 1%). On the other hand, it provided a high AUC value (mean ± standard error: 62 ± 5%) and increased Sensitivity (mean ± standard error: 47 ± 8%) and F1 (mean ± standard error: 29 ± 5%) values. At the same time, it gave Sensitivity values >0 for 17 out of 22 pathological patients. Moreover, the Gaussian SVM model gave higher Sensitivity values for 64% of pathological patients than those given by the Linear SVM.

Concerning Sensitivity and Specificity, our results are slightly worse than those reported by Greene et al. [Greene et al., 2007b] and Doyle et al. [Doyle et al., 2010]. However, a comparison is challenging because a standardized framework for performance assessment for the seizure detection task is currently missing and the metrics used to report NSD systems results vary in the literature [Olmi et al., 2021]. Moreover, our study was retrospective as we trained and validated our methods on ECGs collected by the medical staff in NICUs, with no artefact correction to improve the recording quality. Thus, raw signals could be affected by noisy artefacts such as those due to natural newborns' motor activity and therapeutic manoeuvres performed by the clinicians. Furthermore, most NSD systems proposed in the literature are evaluated on private datasets only. Greene et al. [Greene et al., 2007b] proposed a Linear SVM for automatic HR-based seizure detection based on a dataset of eight ECG recordings from seven full-term newborns admitted in NICU for HIE. This dataset contained 101 h of recordings. Doyle et al. [Doyle et al., 2010] proposed a Linear SVM that was evaluated on a dataset of 208 h of recordings from 14 full-term newborns admitted in NICU for HIE.

In this study, we considered a dataset of about 46 h of recordings from CD. Even though the dataset is made of fewer hours of recordings with respect to [Greene et al., 2007b] and [Doyle et al., 2010], it concerns a wider set of subjects which is made of 51 full-term babies, 22 of which have seizures, and 29 are control patients. To the best of our knowledge, our study is the first one that proposes an ECG-based NSD system focusing the analysis also on a set of control newborns, thus offering a more representative picture of the performance of the system in a real clinical environment. Indeed, considering also healthy newborns would provide an effective NSD system to support clinicians in distinguishing between pathological and healthy patients.

Overall, the performances of the ECG-based systems are lower than those provided by the EEG-based systems [Olmi et al., 2021]; thus, they seem not suitable for clinical implementation. EEG is the basic technique for detecting neonatal seizures as it allows recording and analyzing the spontaneous electrical cerebral activity [Pressler et al., 2021]. The reasons behind this gap in performance between EEG-based NSD and ECG-based systems could be several and heterogeneous [Olmi et al., 2021, Statello et al., 2021, Doyle et al., 2010]. However, there is still room for improvement in HRV features for neonatal seizure detection. For example, searching for different and specific nonlinear features that can better describe the heart rate dynamics in neonatal patients could be advisable [Frassineti et al., 2021d, Lucchini et al., 2016]. Furthermore, there is still a lack of information about possible relationships between the Autonomic Nervous System and the Central Nervous System during neonatal seizures. Thus, understanding and characterizing the interactions between the two systems during or close to ictal events in the newborn might add helpful information in the seizure detection problem [Valenza et al., 2016, Frassineti et al., 2022a]. Moreover, to the best of our knowledge, for ECG/HRV analysis, no deep-learning method was proposed in the literature for neonatal seizure

detection. Considering the improvement obtained by DL techniques on EEG, these methods should also be evaluated on NSD experiments with ECG signals [Olmi et al., 2021].

In conclusion, the works concerning the possibility to develop HRV-based NSD are reported from section 5.1 to 5.3. Results are promising, although with a lower performance with respect to EEG-based both for GLM methods (section 5.2) and SVM methods (section 5.3), while the Patient Discriminant approaches (section 5.1) are comparable to EEG-based methods [Frassineti et al., 2021b]. Thus, the results suggested that HRV analysis and AI models could be included as support tool to clinical staff for neonatal seizure detection and characterization.

5.4 A MATLAB tool for NSD

In this section a short introduction of the MATLAB tool for NSD developed during this PhD is reported. The MATLAB tool implements some of the methods proposed from Chapter 2 to 5 in order to support the clinical staff in the neonatal seizure detection and characterization tasks. In figure 5.6 a screenshot of a preliminary version of the developed interface is provided. The first alpha version will be released.

In particular, the following approaches were implemented:

- Patient Discriminant by HRV or EEG features (section 5.1) [Frassineti et al., 2021b]
- EEG-based NSD, using multichannel EEG features (Chapter 4) [Frassineti et al., 2020, Frassineti et al., 2021a]
- ECG-based NSD (sections 5.2 and 5.3) [Frassineti et al., 2021c, Olmi et al., 2022a]

In addition to the methods discussed in this PhD thesis, the MATLAB tool implements other approaches:

- Video analysis of the newborn's face, extracting features for the seizure detection [Olmi et al., 2022b].
- Aetiology characterization (HIE vs. non-HIE) by EEG and ECG features [Frassineti et al., 2022b].

Furthermore, the application allows to extract all the EEG/ECG features (Figure 5.6, button "Extract Features") calculated during the implicit sub-windowing procedure. The tool is also capable to produce offline reports for the clinical staff [Frassineti et al., 2019], highlighting the part of the exam classified as ictal by the AI models and generating text reports that summarize all the essential information for the neurological assessment of the newborn: number of seizure events detected, time occurrence of the seizure events, EEG derivations involved, maximum and average seizure duration etc..

This MATLAB tool was developed in collaboration with the Neurophysiology Unit from the AOU Careggi University. Moreover, it is currently under development, thus both the interface and its functions may change in future developments. In the next future we are going to validate it in the clinical practice, following the approach proposed by Malak et al. [Malak et al., 2018]. Therefore, the proposed AI methods will be tested on other datasets than the Helsinki and the Careggi ones, evaluating the performance both on retrospective data and prospective data, in order to evaluate if the use of NSD could be helpful to the diagnosis and the treatment of neonatal seizures.

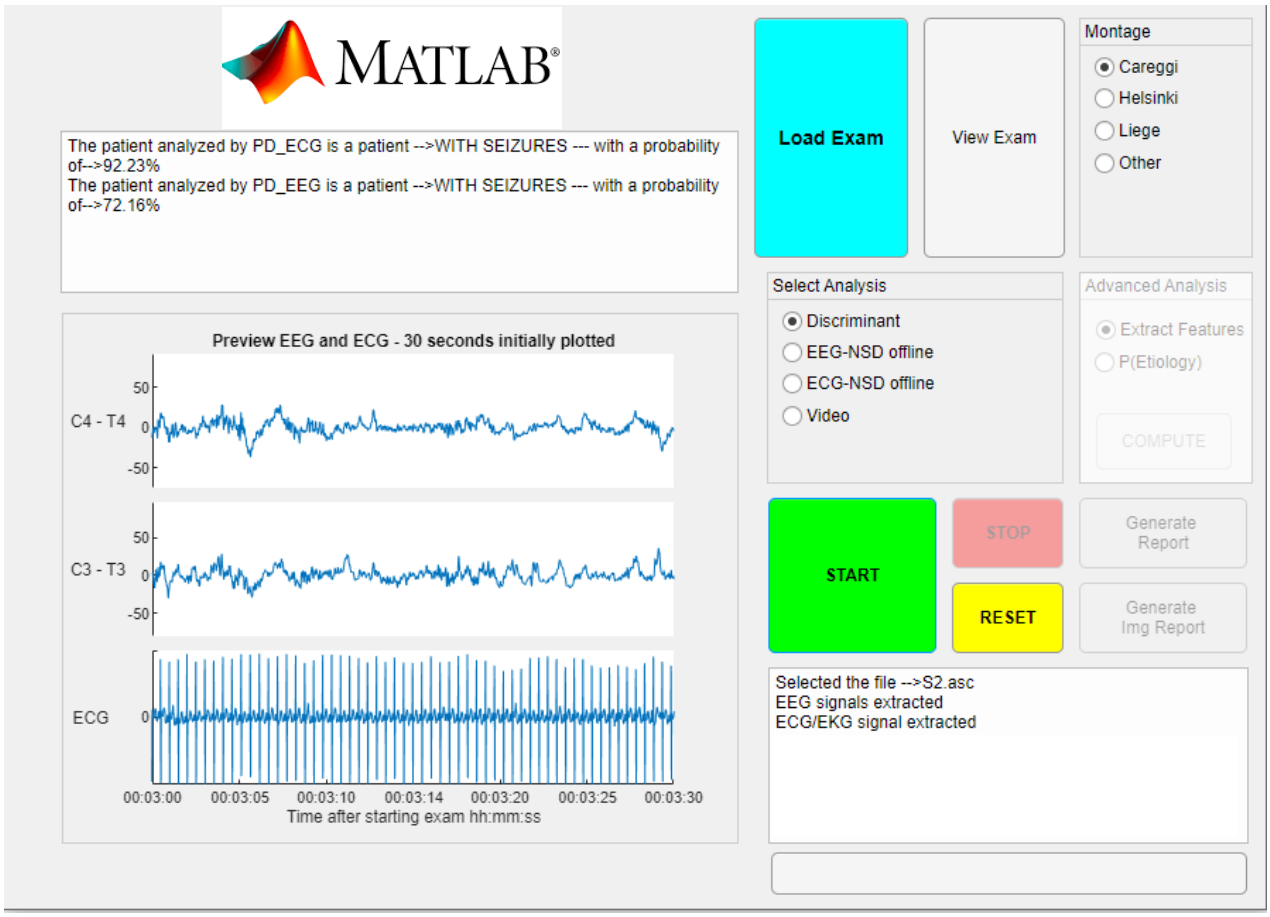


Figure 5.6. Screenshot of the developed interface.

6. Quantitative acoustical analysis in genetic syndromes: towards the definition of a speech phenotype

Some contents in this chapter are based on the following publications:

- *Frassineti, L., Zucconi, A., Calà, F., Sforza, E., Onesimo, R., Leoni, C., Rigante, M., Manfredi, C., Zampino, G., 2021 . Analysis of vocal patterns as a diagnostic tool in patients with genetic syndromes. In Models and analysis of vocal emissions for biomedical applications: 12th international workshop: December 14-16, 2021: Firenze, Italy (pp. 83-86). doi: http://digital.casalini.it/9788855184496.*
- *Frassineti, L., Calà, F., Sforza, E., Onesimo, R., Leoni, C., Lanatà, A., Zampino, G., Manfredi, C., (2022 under review). Quantitative acoustical analysis of genetic syndromes in the number listing task. Biomedical Signal Processing and Control.*

Moreover, some contents in this chapter are based on the following Master's Degree Thesis:

- *Calà, Federico. Thesis title: AI techniques for the acoustic characterization of genetic syndromes. Università degli Studi di Firenze (Firenze, Italy), Scuola di Ingegneria, Curricula Biomedical Engineering, date of discussion: 04/11/2022. Supervisors: Prof. Claudia Manfredi, Prof. Antonio Lanatà. Co-Supervisor: Lorenzo Frassineti. [Calà, 2022].*

In the previous chapters of this PhD thesis an extensive analysis regarding quantitative methods to detect and characterize seizures in children and newborns has been presented. It was shown how Artificial Intelligence (AI) methods and quantitative analysis of physiological signals can provide helpful tools, supporting the clinical staff in the diagnosis of several neurological disorders related to seizures. Thus, AI techniques may have several applications for paediatric subjects in the practice.

In this chapter another application of AI methods and quantitative analysis of biosignals is presented. Specifically, the analysis of voice recordings of subjects with genetic syndrome is addressed. This evaluation was performed towards the definition of a speech phenotype for genetic syndromes through a completely non-invasive procedure. The analysis is based on voice recordings obtained by smartphones or microphones, both in hospital setting and at home. This research was carried on taking into account two main aims: defining quantitative methods for the characterization of a speech phenotype in genetic syndromes and providing methodologies for monitoring over time of the voice characteristics of the subject.

This chapter is organized as follows: a statistical analysis is carried on regarding the main acoustical features of some genetic syndromes obtained during a specific speech task: the number listing task. Then, AI methods are applied to discriminate among genetic syndromes with different speech phenotypes.

6.1 Introduction

In recent years, acoustical analysis has been increasingly applied as a non-invasive tool to characterize genetic syndromes. Anomalies in voice quality have been detected in several genetic syndromes, providing evidence that a specific speech phenotype could exist for some of them [Stojanovik, 2021]. Therefore, acoustical analysis could provide additional information for the characterization of the diseases and in the follow-up of the vocal capabilities of the subject [Hidalgo-De la Guía et al., 2021a, Calà, 2022]. However, to the best of our knowledge, only a few works applied quantitative acoustical

analysis along with qualitative voice assessment in the study of genetic syndromes [Stojanovik, 2021, Hidalgo-De la Guía et al., 2021a, Lazzaro et al., 2020, Corrales-Astorgano et al., 2018]. Thus, quantitative voice analysis could improve the characterization of language phenotyping [Calà, 2022].

Among genetic syndromes, Down syndrome (DS) is one of the most studied as far as its voice/acoustical properties are concerned [Moura et al., 2008]. In about 90% of cases, DS comes from an extra copy of chromosome 21 (Trisomy 21) and is characterized by a distinctive facial appearance, intellectual disability and muscular hypotonia [Antonarakis et al., 2020]. Common findings include alterations of the vocal tract structure including ogival palate, pharynx constriction, arytenoids cartilages thickening, laryngomalacia, and macroglossia. The reduced vocal tract volume results in speech impairments and limited articulation [Moura et al., 2008, Chin et al., 2014]. In adults, recent MRI studies underlined a smaller volume of frontal and temporal lobes, where Broca and Wernicke areas respectively placed [Hamner et al., 2018]. Several works proved the usefulness of the analysis of quantitative vocal features to better characterize the syndrome and as a support to monitor the voice characteristics of the DS subjects over time [Corrales-Astorgano et al., 2018, Moura et al., 2008], through acoustical analysis with applications in rehabilitation and logopaedics. Several research findings confirmed differences between DS and control subject regarding disfluency problems (stuttering or cluttering) and poor control over energy in stressed versus unstressed vowels [Lazzaro et al., 2020]. Moreover, several biomechanical studies confirmed the differences in vocal production for DS subjects, confirming that this information could be used to define a specific language phenotype for these subjects [Hidalgo-De la Guía et al., 2021b].

Regarding voice disorders, increasing evidence was found for subjects affected by the Noonan Syndrome (NS; OMIM #163950) [Lazzaro et al., 2020]. NS is a genetically inherited disease due to gene mutations involving the RAAS/MAPK (mitogen-activated protein kinase) signalling pathway [Roberts, 2001]. Classified as a RASopathy, it is characterized, among other features, by remarkable facial features, including micrognathia, dental crowding, ogival palate. Growth delay, cardiac abnormalities, learning impairment, verbal and non verbal skills difficulties were also reported [Myers et al., 2014]. As stated by Lazzaro et al. [Lazzaro et al., 2020], most of the NS subjects may show specific voice characteristics. Thus, these findings suggest the use of quantitative acoustical analysis to better characterize NS subjects.

Another remarkable example of quantitative acoustical analysis in genetic syndromes is the Smith-Magenis Syndrome (SMS; OMIM #182290). SMS is a rare genetic syndrome frequently caused by a heterozygous deletion of or a heterozygous pathogenic variant in *RAI1* on chromosome 17p11.2 [Hidalgo-De la Guía et al., 2021a]. SMS subjects present mental retardation, behavioural abnormalities, sleep disorders, and early onset obesity [Gropman et al., 2006]. Common dysmorphisms include midface retrusion, short and broad nose, everted and tented upper lip vermilion, prognathism. Otolaryngological manifestations include velopharyngeal insufficiency, high tendency to nodules and polyps' formation, vocal cords oedemas, and paralysis [Gropman et al., 2006]. Moreover, as stated by Hidalgo-De la Guía et al. [Hidalgo-De la Guía et al., 2020], SMS subjects show several peculiar vocal and biomechanical characteristics, making the syndrome a promising candidate for language phenotyping.

According to the existing literature, the acoustical analysis applied to genetic syndromes mainly focused on the evaluation of time-frequency properties of repeated and sustained vowels [Hidalgo-De la Guía et al., 2021a, Corrales-Astorgano et al., 2018]. In particular for Italian speakers, the most studied vowels are /*a*/, /*i*/, and /*u*/ (that roughly correspond to "a", "i", "u" in the International Phonetic Alphabet [Deller et al., 2000]), that are quite stable against dialectal inflections. More

complex tasks, such as the list of the sequence of the months, were already evaluated for some neurodegenerative diseases [König et al., 2015, Englert et al., 2019, Bandini et al., 2015]: this was found helpful to better understand both the cognitive and the vocal skills of the subjects.

However, to the best of our knowledge, only a few works concern the quantitative acoustical and speech properties in genetic syndromes using tasks different from sustained vowels [Stojanovik, 2021, Corrales-Astorgano et al., 2018]. Indeed, the analysis of more complex tasks, such as the list of numbers or structured texts, has not been fully exploited yet. Furthermore, the analysis of multi-domain features might provide complementary information about the language phenotype for a specific syndrome, such as features from vowels or more complex speech tasks. Acoustical features could also be used with machine-learning models for the characterization of genetic syndromes [Frassineti et al., 2021e].

In this work, we evaluated if the acoustical analysis of a task more complex than the repetition of sustained vowels may provide helpful additional information to characterize three different genetic syndromes: DS, NS and SMS, as compared both to control subjects and among the syndromes, to highlight possible inter-syndrome differences. In particular, the task of listing numbers in ascending order from 1 to 10 in Italian is exploited here. To this aim, statistical analysis was applied to several acoustical features extracted with the BioVoice tool [Morelli et al., 2021]. Furthermore, a multiscale sample entropy-based approach was added to the acoustical features to assess if such method could discriminate between pathological and control voices, on analogy to [Arias-Londoño et al., 2010, Mekyska et al., 2015]. Both the whole number listing task and each single number were analysed to find possible differences among subjects.

6.2 Material and methods for quantitative acoustical analysis

All participants were recruited among those routinely monitored at the Rare Disease Unit of Paediatrics Department, Fondazione Policlinico Agostino Gemelli-IRCCS, Rome, Italy, over a two years period [Calà, 2022]. Signed informed consent was provided by parents/caregivers. All patients were clinically and genetically characterized. Specifically, the studied population included: 24 subjects with Down syndrome (DS), 24 with Noonan syndrome (NS), 24 with Smith-Magenis syndrome (SMS), and 15 Control Subjects (CS). The age range was 4-18 years, with a mean of 10.8 ± 3.7 years. The Kruskal-Wallis test found no statistical differences concerning the age among the four groups (level of significance 0.05). In Table 6.1 details about age and gender are reported. The acronym PA (Paediatric Age) identifies the subjects less than 12 years old.

The recordings were performed according to a study based on the SIFEL protocol [Ricci and Maccarini, 2002], the Italian version of the European Laryngological Society (ELS) protocol [Dejonckere et al., 2001] that aims at providing a functional assessment of voice pathology. The recorded audio files consist of the following tasks:

1. Italian vowels /a/, /i/, /u/, /ɔ/ and /ɛ/ sustained for at least 4 seconds.
2. The Italian word “aiuole” (/aˈjwɔle/, flower beds).
3. the list of the Italian numbers from one to ten, in ascending order: “uno” (/’uno/); “due” (/’due/); “tre” (/’tre/); “quattro” (/’kwattro/); “cinque” (/’ʧinkwe/); “sei” (/’sɛj/); “sette” (/’sɛtte/); “otto” (/’ɔtto/); “nove” (/’nɔve/); “dieci” (/’djɛʧi/). Note that in Italian the numbers 4,7 and 8 contain the “double t” sound: /tt/.

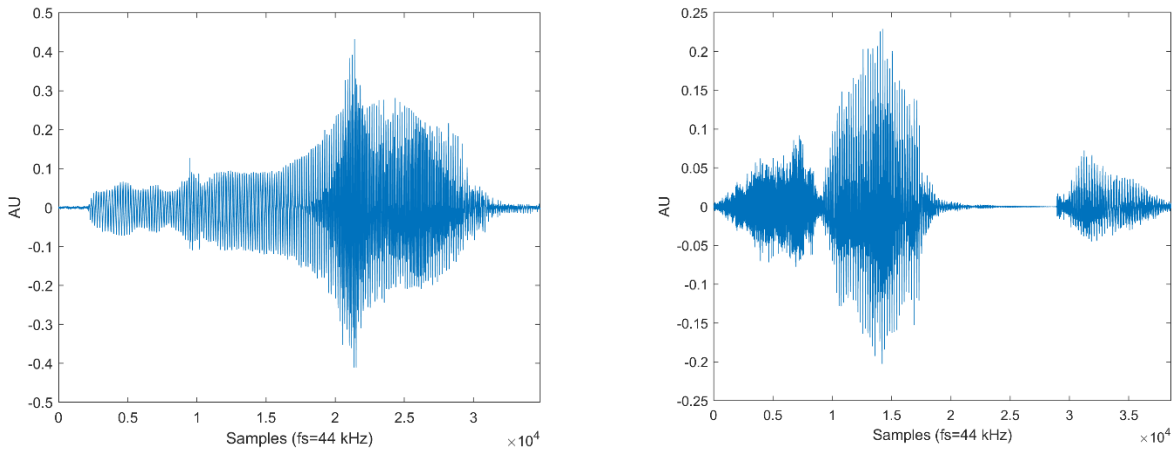
Only the list of numbers (task n.3) was considered in this work. In another study [Frassineti et al., 2021e], the vowels (task n.1) were already analysed, showing differences among pathological subjects. Table 6.1 shows the dataset. The number of PA subjects is reported in round brackets, the acronym *iqr*, reported in brackets in the column "Age", refers to the interquartile range. As shown in Table 6.1, despite the possible cognitive delay of the subjects, almost all of them correctly emitted the whole numerical sequence. Only a small subset of DS subjects (5 subjects) and 1 SMS did not complete the entire task. Recordings were obtained using a commercial smartphone in a controlled environment (environmental noise <40dB), with the smartphone set at 15 centimetres from the subject's lips and with an angle of 45°. The sampling rate was 44100Hz. All the analyses were performed using the BioVoice tool [Morelli et al., 2021] and routines implemented in MATLAB 2021b. BioVoice is a voice analysis tool freely downloadable at https://github.com/ClaudiaManfredi/BioVoice.

Table 6.1 – Subjects involved in the study. PA subjects are reported in brackets in the first and last column. [Frassineti et al., 2022 under review].

Cases	**Number of subjects (PA)**	**Age** ***Median (iqr)***	**Gender** ***Male/Female***	**Subjects unable to perform the whole task (PA)**
CS	15 (9)	11 (5)	7/8	0
DS	24 (17)	10 (4.5)	16/8	5 (4)
NS	24 (13)	12 (7.5)	11/13	0
SMS	24 (13)	11 (6)	7/17	1 (0)

6.2.1 Manual Annotation and Automatic Voiced/Unvoiced Detection

Before starting the quantitative acoustical analysis, each recording was manually annotated, labelling the onset and offset of each number emitted by the subjects. In Figure 6.1, the temporal profile of two numbers "due" and "sette" (2 /'due/ and 7 /'sɛtte/) from the same control subject are shown. This example highlights differences between an almost "vocalic number" (Figure 6.1a "due" /'due/) and a number with the double /t/ inside (Figure 6.1b "sette", /'sɛtte/). Indeed, numbers with /tt/ are characterized by two "voiced" parts and an unvoiced part in the middle corresponding to /tt/.



(a) (b)

Figure 6.1. Temporal profile of two recorded numbers from a control subject. (a) = /'due/ (2); (b) = /'sɛtte/ (7). [Frassineti et al., 2022 under review].

Based on the manual annotations, the following parameters were estimated:

- **average time between each number** (aTBN), defined as the average time difference in seconds between the offset and the onset of two consecutive numbers [König et al., 2015];
- **standard deviation time between each number** (sTBN);
- **voice segment length** (VSL), i.e. the temporal length in seconds of each number in the sequence (between its onset and offset) [König et al., 2015, Bandini et al., 2015];
- **percentage of unvoiced segments** (PUVS), defined as the percentage of the whole recording without voiced segments (starting from the offset of the first number to the onset of the last number) [Bandini et al., 2015];
- **task length**, defined as the time difference in seconds between the onset of the first number and the offset of the last number, including also unvoiced segments.

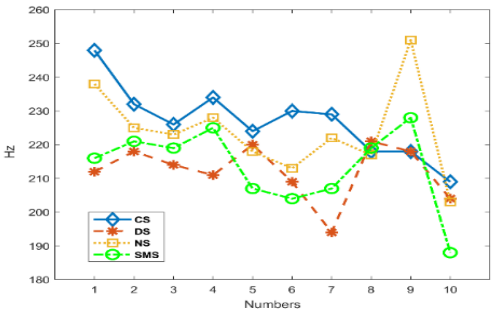
These parameters were selected according to previous studies [König et al., 2015, Englert et al., 2019, Bandini et al., 2015], as they were found helpful for voice characterization in some diseases, such as Alzheimer's or Parkinson's [König et al., 2015, Bandini et al., 2015]. Furthermore, as in [Bandini et al., 2015], the recordings were analyzed with two different voiced/unvoiced detectors. One is implemented in BioVoice and was already used in a previous study to evaluate speech anomalies in subjects affected by Parkinson disease [Bandini et al., 2015]. The second one was the built-in speech detector available in the MATLAB tool (function detectSpeech version 2021b).

Further details are reported in [Giannakopoulos, 2009]. In this work, the number of voiced segments found by the two detectors was compared to the manual labels, that make up the ground truth. In particular, the absolute error between the ground truth and the detected voiced segments was considered: the absolute error is equal to 0 if the number of detected segment match the manual annotations. Thus, for example, 10 voiced segments expected and 11 voiced segments detected correspond to an absolute error equal to 1. Results concerning this analysis are reported in Section 6.3.

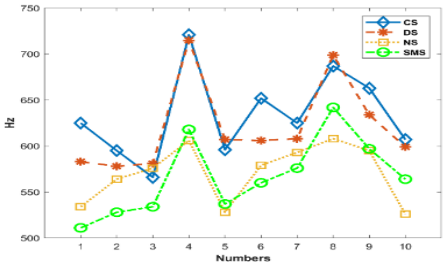
6.2.2 BioVoice and Multiscale Entropy Acoustical Analysis

With BioVoice, 30 acoustical features were extracted. The analysis was performed distinguishing between paediatric subjects (<12 years old) and adults [Deller et al., 2000] and, in the case of adults, between males and females. According to [Deller et al., 2000, Morelli et al., 2021], the cut-off of 12 years of age was added to provide a more reliable evaluation of acoustical parameters with BioVoice. The 30 acoustical parameters estimated with BioVoice are: maximum, minimum, mean, median, and standard deviation for F0 and formants F1, F2, and F3; $T0_{min}$ and $T0_{max}$ for F0; jitter; Normalized Noise Energy (NNE); signal and voiced part duration; mean, min and max of voiced segments; mean of pause (unvoiced) duration.

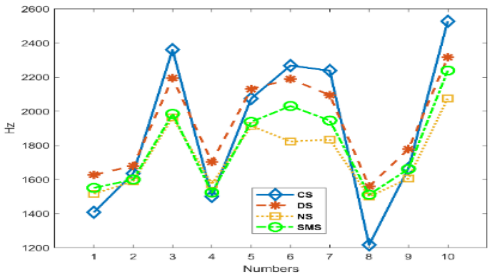
These features are in line with existing literature and were proven to be correlated with several physiological mechanisms linked to voice production [Deller et al., 2000]. Further details about such acoustical features are reported in [Morelli et al., 2021]. For all the cases considered, the estimated values of F0 and formants F1 and F2 are reported in Table 6.2, and 6.3, respectively. Figure 6.2 shows the frequency median trends for the three acoustical parameters (F0, F1, and F2) for the four groups considered and numbers 1-10.



(a) F0



(b) F1



(c) F2

Figure 6.2. Frequency median trends of numbers 1-10 for: (a) F0, (b) F1, (c) F2. CS: solid line(◊); DS: dashed line (*) ;NS: dotted line (□): SMS: dash-dotted line (○). [Frassineti et al., 2022 under review].

Table 6.2. Median and iqr values of F0 for each group and for all the numbers in the task. CS=Control Subjects (15 cases), DS= Down Syndrome, NS=Noonan Syndrome and SMS= Smith-Magenis Syndrome (24 cases for each syndrome). [Frassineti et al., 2022 under review].

	F0 [Hz] Median (iqr)			
Number	**CS**	**DS**	**NS**	**SMS**
1	248 (72)	212 (102)	238 (58)	216 (51)
2	232 (68)	218 (66)	225 (55)	221 (55)
3	226 (85)	214 (81)	223 (46)	219 (44)
4	234 (93)	211 (82)	228 (65)	225 (45)
5	224 (66)	220 (72)	218 (35)	207 (57)
6	230 (83)	209 (100)	213 (56)	204 (54)
7	229 (81)	194 (59)	222 (51)	207 (44)
8	218 (79)	221 (108)	217 (71)	219 (65)
9	218 (94)	218 (68)	251 (51)	228 (85)
10	209 (77)	204 (109)	203 (55)	188 (56)

Table 6.3. Median and iqr values of formants F1 and F2 for each group and for all the numbers in the task. CS=control subjects (15 cases), DS= Down Syndrome, NS=Noonan Syndrome and SMS= Smith-Magenis Syndrome (24 cases for each syndrome). [Frassineti et al., 2022 under review].

	F1 [Hz] Median (iqr)				**F2 [Hz] Median (iqr)**			
Number	**CS**	**DS**	**NS**	**SMS**	**CS**	**DS**	**NS**	**SMS**
1	625 (226)	583 (206)	534 (304)	511 (259)	1408 (344)	1628 (522)	1516 (257)	1551 (353)
2	595 (252)	578 (199)	564 (264)	528 (231)	1636 (220)	1680 (343)	1590 (287)	1600 (316)
3	566 (110)	581 (167)	576 (306)	534 (232)	2361 (440)	2195 (470)	1964 (394)	1985 (447)
4	721 (128)	715 (249)	606 (323)	618 (244)	1501 (325)	1703 (434)	1573 (279)	1524 (299)
5	596 (173)	607 (211)	528 (236)	537 (241)	2075 (616)	2130 (603)	1914 (480)	1938 (515)
6	652 (215)	606 (220)	579 (258)	560 (224)	2268 (761)	2189 (509)	1823 (507)	2031 (440)
7	625 (146)	608 (188)	593 (305)	576 (205)	2238 (641)	2095 (469)	1834 (293)	1946 (300)
8	687 (214)	699 (240)	608 (349)	642 (221)	1217 (313)	1560 (327)	1500 (237)	1511 (282)
9	663 (196)	634 (176)	595 (296)	597 (254)	1665 (389)	1778 (477)	1606 (272)	1662 (289)
10	607 (195)	599 (221)	526 (193)	564 (247)	2528 (768)	2316 (780)	2076 (572)	2239 (553)

Besides the acoustical features, a preliminary evaluation based on multiscale entropy analysis was performed. In acoustical analysis, entropy indexes were already used to discriminate between pathological and healthy voices [Arias-Londoño et al., 2010, Fontes et al., 2014, Sun et al., 2017]. The multiscale sample entropy index (MSE) was used in [Costa et al., 2005].

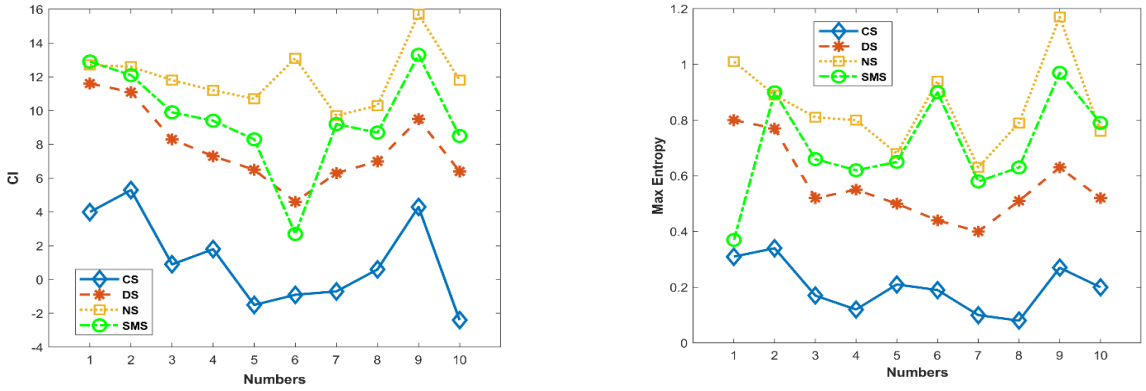
Recently, the multi-scale approach applied to physiological signals was motivated by successful findings in detecting more helpful information rather than the single scale approach [Frassineti et al., 2021d]. Sample Entropy (SE), already introduced in Chapter 4 of this PhD Thesis, is one of the most used entropy measures for the analysis of physiological signals [Richman and Moorman, 2000].

Low SE values are generally related to more predictable and regular time series [Humeau-Heurtier, 2015]. According to Costa et al. [Costa et al., 2005], the parameters for estimating MSE values are: embedding dimension m=3, threshold r=0.2, and number of scales =20. The coarse-grained procedure, described in Equation 4.1, was used to generate the scales [Costa et al., 2005, Frassineti et al., 2021d]. From the 20 values of MSE, the following measures were extracted:

- the complexity index (**CI MSE**) [Costa et al., 2005, De Wel et al., 2017].
- the maximum entropy (**Max MSE**), defined as the maximum value of MSE among the 20 scales considered. [Costa et al., 2005, De Wel et al., 2017].

These two measures were already used to describe complexity properties of physiological signals by multiscale entropy approaches [Frassineti et al., 2021d, Humeau-Heurtier, 2015, De Wel et al., 2017]. Values of CI MSE and Max MSE are reported in Table 6.4 and Table 6.5, respectively.

In Figure 6.3 the median trends for the two multiscale entropy indexes (CI and Max MSE) are shown for the four groups considered and numbers 1-10.



(a) CI MSE (b) Max MSE

Figure 6.3. Median trends of numbers 1-10 for multiscale entropy indexes: (a) CI MSE, (b) Max MSE. CS: solid line(◊); DS: dashed line (*); NS: dotted line (□); SMS: dash-dotted line (○). [Frassineti et al., 2022 under review].

Table 6.4. Values (median and iqr) of MSE Complexity Index (using 20 scales) for all the groups considered: CS=control subjects (15 cases), DS= Down Syndrome, NS=Noonan Syndrome and SMS= Smith-Magenis Syndrome (24 cases for each syndrome). [Frassineti et al., 2022 under review].

	MSE Complexity Index (scales=20) All subjects Median (iqr)			
Number	**CS**	**DS**	**NS**	**SMS**
1	4.0 (8.1)	11.6 (3.5)	12.7 (5.5)	12.9 (4.2)
2	5.3 (2.7)	11.1 (4.2)	12.6 (3.1)	12.1 (4.6)
3	0.9 (6.5)	8.3 (7.9)	11.8 (9.2)	9.9 (9.6)
4	1.8 (1.9)	7.3 (5.1)	11.2 (6.9)	9.4 (6.8)
5	-1.5 (5.7)	6.5 (7.4)	10.7 (7.6)	8.3 (7.2)
6	-0.9 (4.5)	4.6 (12.9)	13.1 (23)	2.7 (24)
7	-0.7 (3.3)	6.3 (5.1)	9.7 (8)	9.2 (5.8)
8	0.6 (2.5)	7.0 (7.3)	10.3 (5.6)	8.7 (6.5)
9	4.3 (4.9)	9.5 (3.7)	15.7 (9.2)	13.3 (7.8)
10	-2.4 (3.5)	6.4 (11.4)	11.8 (6.2)	8.5 (20.9)

Table 6.5. Values (median and iqr) of Max Entropy of MSE (using 20 scales) for all the groups considered: CS=control subjects (15 cases), DS= Down Syndrome, NS=Noonan Syndrome and SMS= Smith-Magenis Syndrome (24 cases for each syndrome). [Frassineti et al., 2022 under review].

	Max Entropy MSE (scales=20) All subjects Median (iqr)			
Number	**CS**	**DS**	**NS**	**SMS**
1	0.31 (0.23)	0.80 (0.38)	1.01 (0.45)	0.37 (0.11)
2	0.34 (0.21)	0.77 (0.33)	0.89 (0.18)	0.90 (0.36)
3	0.17 (0.16)	0.52 (0.43)	0.81 (0.70)	0.66 (0.54)
4	0.12 (0.08)	0.55 (0.34)	0.80 (0.53)	0.62 (0.50)
5	0.21 (0.12)	0.50 (0.45)	0.68 (0.47)	0.65 (0.48)
6	0.19 (0.22)	0.44 (0.38)	0.94 (0.52)	0.90 (0.51)
7	0.10 (0.08)	0.40 (0.25)	0.63 (0.53)	0.58 (0.46)
8	0.08 (0.04)	0.51 (0.44)	0.79 (0.52)	0.63 (0.57)
9	0.27 (0.27)	0.63 (0.30)	1.17 (0.70)	0.97 (0.66)
10	0.20 (0.24)	0.52 (0.29)	0.76 (0.32)	0.79 (0.39)

6.2.3. Statistical Analysis

This study aims at evaluating if the measurements obtained from the manual annotations defined in section 6.2.1 (e.g. aTBN, absolute errors of voiced/unvoiced detectors etc…), the acoustical features extracted by BioVoice and the entropy indexes obtained with MATLAB® (version 2021b) routines allow discriminating between subjects with genetic syndromes and control subjects. Moreover, it was evaluated if the same features were able to discriminate among the syndromes, thus finding possible inter-syndrome differences.

First, the hypothesis of normality distribution was checked with the Shapiro-Wilk test (level of significance p=0.05) for all the parameters considered, distinguishing among groups. If the normality hypothesis was rejected, the non-parametric Kruskal-Wallis test was applied, otherwise the ANOVA test was performed on the data. For both tests, a level of significance p=0.05 was considered. Then, a Dunn-Sidak multiple comparison post hoc correction [Dinno, 2015] was applied in order to evaluate pairwise differences (e.g. DS vs. CS or inter-syndrome such as DS vs. NS).

Regarding the acoustical features, all the statistical tests were evaluated for each number of the task. Furthermore, a stratified statistical analysis was included, considering only the subjects less than 12

years old for all the groups. Indeed, with BioVoice the acoustical features are estimated distinguishing between children and adults. Therefore, this analysis was added to assess if the parameters obtained from adults might introduce confounding effects in the statistical results.

6.3 Results

In this section, the statistical results obtained for each acoustical parameter are reported. Table 6.6 shows the descriptive statistics (median and iqr) concerning the parameters derived from the manual annotations described in Section 6.2.1. The statistically significant parameters between control subjects and those among genetic syndromes are reported in bold. Regarding inter-syndrome evaluations, "δ" denotes significant differences between DS than SMS and NS, and "α" denotes significant differences between DS and NS. In Table 6.7 the statistical results are reported concerning the absolute error between the manual annotations and the two automatic voiced/unvoiced detectors (BioVoice and SpeechDetect in MATLAB). The statistically significant parameters between syndromes and CS are reported in bold. For the SMS subjects a significant difference between the two tools in terms of absolute error was found.

Table 6.6. Results of statistical analysis on qualitative acoustical features for controls subjects and syndromes. The statistically significant parameter for each syndrome with respect to control subjects are shown in "bold". "**δ**" denotes parameters that show significant differences for DS with respect to the other two syndromes (SMS and NS). "**α**" denotes significant differences between DS and NS. [s]=seconds. [Frassineti et al., 2022 under review].

	Syndrome Median (iqr)			**CS Median (iqr)**
Parameter	**SMS**	**NS**	**DS**	
aTBN [s]	0.25 (0.33)	**0.24 (0.16)**	0.38 (0.27)$^{\delta}$	0.43 (0.30)
sTBN [s]	0.09 (0.08)	0.10 (0.04)	**0.17 (0.13)**$^{\delta}$	0.11 (0.04)
VSL [s]	**0.52 (0.17)**	**0.46 (0.08)**	0.60 (0.22)$^{\delta}$	0.70 (0.20)
PUVS [%]	26 (18)	28 (12)	33 (11)	29 (15)
task length [s]	**7.50 (3.61)**	**6.36 (2.18)**	9.09 (3.73)$^{\alpha}$	10.08 (4.45)

Table 6.7. Results for "voiced" detection experiment using two different tools: BioVoice and SpeechDetect from MATLAB. The absolute errors (median and iqr) between the voiced segments detected by the tools and the ground truth are shown. Bold= statistical difference between control subjects and the subjects with genetic syndrome. "**α**" denotes significant differences between the errors obtained by the two tools on any group (control or pathological). [Frassineti et al., 2022 under review].

Absolute error	**Syndrome Median (iqr)**			
Tool	**DS**	**NS**	**SMS**$^{\alpha}$	**CS**
BioVoice	**4.5 (6.5)**	**6.5 (7)**	**9 (5)**	3 (2.5)
MATLAB – SpeechDetect	3 (4.5)	**7.5 (6.25)**	**6 (8.25)**	2 (2)

In Table 6.8 the number of significant acoustical features found for each number are reported. Both the global number (Significant features vs. Controls), the pairwise differences and the inter-syndrome differences are reported. In Table 6.9 the number of significant features obtained considering only PA subjects are shown. Further results about which features were found significant for each syndrome and for both cases (all subjects and PA subjects only) are reported in Tables 6.10, 6.11 and 6.12.

Table 6.8. Number of significant acoustical features found for all the subjects considered. [Frassineti et al., 2022 under review].

	All subjects Number of significant acoustic features				
Number	***Significant features vs. CS***	***DS vs. CS***	***NS vs. CS***	***SMS vs. CS***	***Inter-syndromes (DS vs. NS; DS vs. SMS; NS vs. SMS)***
1	9	8	5	5	2
2	6	4	6	4	2
3	6	2	4	4	3
4	10	10	4	6	0
5	7	4	5	5	0
6	6	3	4	6	2
7	11	10	10	9	3
8	14	13	8	14	1
9	6	4	4	4	4
10	3	0	2	2	3

Table 6.9. Number of significant acoustical features found for PA subjects. [Frassineti et al., 2022 under review].

	PA subjects Number of significant acoustic features				
Number Repeated	***Significant features vs. CS***	***DS vs. CS***	***NS vs. CS***	***SMS vs. CS***	***Inter-syndromes (DS vs. NS; DS vs. SMS; NS vs. SMS)***
1	8	4	2	8	2
2	9	4	6	7	0
3	12	5	12	7	4
4	12	11	3	8	1
5	8	5	5	6	1
6	9	4	5	7	0
7	17	14	14	13	4
8	15	15	5	12	8
9	8	2	4	7	5
10	5	1	4	5	1

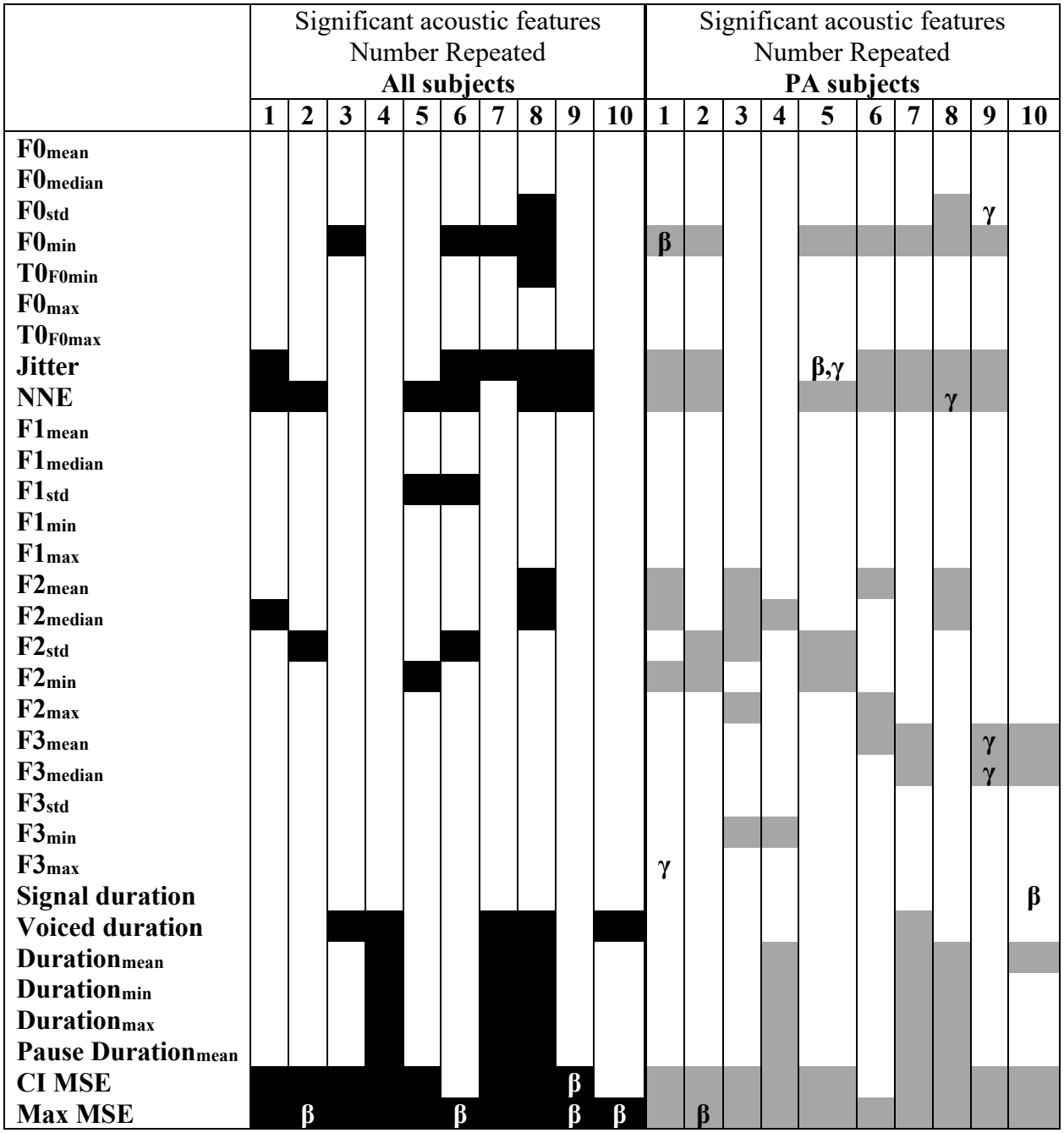
Table 6.10. List of significant acoustical features for all the DS subjects and for the PA subjects subset for all the numbers analysed. In black and in grey the significant parameters are highlighted, after post-hoc correction, as compared to control subjects. “α” denotes an inter-syndrome difference DS vs. NS, “β” denotes an inter-syndrome difference DS vs. SMS. [Frassineti et al., 2022 under review].

	Significant acoustic features Number Repeated **All subjects**										Significant acoustic features Number Repeated **PA subjects**									
Features	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
$F0_{mean}$																	α			
$F0_{median}$																				
$F0_{std}$																				
$F0_{min}$											β			α						
$T0_{F0min}$																				
$F0_{max}$																				
$T0_{F0max}$																				
Jitter															α,β			α		
NNE																				
$F1_{mean}$																				
$F1_{median}$																				
$F1_{std}$																				
$F1_{min}$													α							
$F1_{max}$																				
$F2_{mean}$			α				α						α				α			
$F2_{median}$			α				α						α				α			
$F2_{std}$								α										α		
$F2_{min}$																				
$F2_{max}$																		α		
$F3_{mean}$																				
$F3_{median}$																				
$F3_{std}$																				
$F3_{min}$																		α		
$F3_{max}$																				
Signal duration																				β
Voiced duration																				
$Duration_{mean}$																		α		
$Duration_{min}$																		α		
$Duration_{max}$																		α		
Pause $Duration_{mean}$																				
CI MSE	α						α		α,β	α									α	
Max MSE	α	α,β	α			α,β			α,β	α,β			β				α		α	

Table 6.11. List of significant acoustical features for all the NS subjects and for PA NS subjects for all the numbers analysed. In black and in grey the significant parameters are highlighted as compared to control subjects. "α" denotes an inter-syndrome difference DS vs. NS, "γ" denotes an inter-syndrome difference NS vs. SMS. [Frassineti et al., 2022 under review].

	Significant acoustic features Number Repeated **All subjects**										Significant acoustic features Number Repeated **PA subjects**									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
$F0_{mean}$																	α			
$F0_{median}$																				
$F0_{std}$																				
$F0_{min}$														α					γ	
$T0_{F0min}$																				
$F0_{max}$																				
$T0_{F0max}$																				
Jitter															α,γ			α		
NNE																		γ		
$F1_{mean}$																				
$F1_{median}$																				
$F1_{std}$																				
$F1_{min}$													α							
$F1_{max}$																				
$F2_{mean}$			α				α						α				α			
$F2_{median}$			α				α						α				α			
$F2_{std}$								α										α		
$F2_{min}$																				
$F2_{max}$																		α		
$F3_{mean}$																			γ	
$F3_{median}$																			γ	
$F3_{std}$																				
$F3_{min}$																		α		
$F3_{max}$											γ									
Signal duration																				
Voiced duration																				
$Duration_{mean}$																		α		
$Duration_{min}$																		α		
$Duration_{max}$																		α		
$Pause\ Duration_{mean}$																				
CI MSE	α						α		α	α									α	
Max MSE	α	α	α			α			α	α		β					α		α	

Table 6.12. List of significant acoustical features for all the SMS subjects and for PA SMS subjects for all the numbers analysed. In black and in grey the significant parameters after post-hoc correction are highlighted as compared to control subjects. “β” denotes an inter-syndrome difference DS vs. SMS, “γ” denotes an inter-syndrome difference NS vs. SMS. [Frassineti et al., 2022 under review].



6.4 Discussion

This work evaluates if acoustical and entropy features derived from a speech task in subjects with genetic syndromes may add helpful information for the characterization of possible different speech phenotypes. To our knowledge, this work is one of the first that concerns quantitative features extracted from a speech task rather than sustained vowels only in genetic syndromes.

In this work, the following three genetic syndromes were considered: Down Syndrome (DS), Noonan Syndrome (NS), and Smith-Magenis Syndrome (SMS). They were considered because of the previous evidence in the literature regarding their speech properties [Hidalgo-De la Guía et al., 2021a, Lazzaro et al., 2020, Corrales-Astorgano et al., 2018] and because they are the most numerous groups among the cases collected in the original study [Frassineti et al., 2021e].

Results shown in Table 6.6 suggest that overall differences might exist between control subjects (CS) and subjects affected by the considered genetic syndromes. Moreover, the three syndromes exhibit specific properties that differentiate them from CS.

Specifically, the median value of aTBN parameter of NS (0.24s, iqr 0.16s) significantly differs from the CS group (0.43s, iqr 0.30s), as well as the median value of VSL parameter and task length, showing the tendency for NS to have a faster speech rate than CS. Moreover, DS subjects showed differences for the parameter sTBN as compared to CS: sTBN represents a sort of measure of irregularity of the speech during the number listing task, thus the result shows that the time required between the emission of two consecutive numbers is more variable in DS than in CS (median 0.17s for DS and 0.11s for CS).

Concerning VSL and task length parameters, both SMS and NS subjects present significant differences when compared to CS (median 10.08s for CS, 7.50s for SMS, 6.36s for NS). Therefore, it reasonable to assume that the CS group is more able to optimize articulation and improve intelligibility, resulting in increased VSL and task length parameters.

On the other hand, the DS group showed significant differences when compared to SMS and NS groups, as they show a higher median VSL (0.60s, iqr 0.22s for DS) and task length. In this case, results may primarily be due to neurobiological factor. Specifically, the underlying genetic condition characterized by different levels of cognitive impairment might not allow efficient motor planning of a timely articulation initiation. The scarce tone of the oral-facial muscle [Desai, 1997] and the relatively large tongue as compared to the size of the oral cavity, typically observed in this genetic condition [Guimaraes et al, 2008], could contribute to the higher values of the considered parameters. Moreover, it might be due to the attempt of the DS subjects to mitigate the disfluency during the task (e.g. cluttering and scuttering) [Kenta et al., 2013], or to the general hypotonia and motor-control difficulties in DS subjects [Corrales-Astorgano et al., 2018].

Table 6.1 highlights that only a small number of subjects, mostly PA, did not complete the task, although such syndromes often imply a significant cognitive delay. Thus, the relative ease of the task could encourage its useful application in the acoustical analysis of these syndromes.

Regarding the possibility of automatically detecting the voiced segments by BioVoice or MATLAB SpeechDetect, Table 6.7 shows that the absolute errors were too high for all the syndromes to consider the automatic detection reliable. On the contrary, for CS, the performance was quite good: a median of less than 3 voiced segments was found for the absolute error detection with both tools. This anyway suggests that the absolute error may be a useful parameter to discriminate between pathological and

healthy voices during the number listing task. As shown in Table 6.7, with BioVoice all the errors obtained with genetic syndromes were statistically higher than those obtained with CS.

In this work the quantitative acoustical analysis, both with BioVoice and MATLAB, was performed based on the manual annotations. However, we remark that the possibility to detect voiced segments for these genetic syndromes remains open, as an extensive evaluation of different speech detectors was not addressed in the present work. Moreover, an iterative adaptation of the algorithm used in SpeechDetect was not considered here but it could be addressed in future developments.

The quantitative acoustical analysis reported in Tables 6.8 and 6.9 shows that differences exist between CS and subjects with syndromes for all the numbers 1-10. In fact, at least one acoustical parameter was found significant for each syndrome. Tables 6.10, 6.11 and 6.12 give a complete overview for each syndrome: the main differences were found for the Jitter and NNE parameters for all the syndromes. In the literature, alterations of Jitter and NNE were already proven to be correlated with malformations of the vocal tract, neurological disorders, and SNC abnormalities [Midi et al., 2008, Texeira et al., 2013]. Moreover, several differences were found among features concerning the F1 formant, which is related to structural alterations of the pharynx. Also, significant differences were found among features related to the F2 formant, linked to motor deficits of tongue, lips, and/or jaw [Deller et al., 2000]. With our analysis, the numbers that exhibit higher differences between pathological and control subjects are the numbers 4, 7, and 8 (in Italian: /'kwa**tt**ro/, /'sɛ**tt**e/ and /'ɔ**tt**o): as already pointed out, these numbers present a common characteristic, that is the double /t/. An example of the temporal profile for number 7 is shown in Figure 6.1 for a CS subject. Thus, it is possible to argue that the double /t/ may cause significant alterations in speech for subjects with the genetic syndromes considered here. Indeed, Pierpont et al. [Pierpont et al., 2010], found that NS subjects had difficulty in pronouncing consonant clusters. Tables 6.10, 6.11 and 6.12 show that differences were found for the features relative to "duration": min max, mean and Pause Duration.

In summary, the analysis of the significant acoustical parameters for each syndrome shows that specific differences exist with CS, thus allowing the identification of helpful characteristics for the definition of speech phenotypes. However, as shown in Tables 6.8, 6.9, 6.10, 6.11 and 6.12, only a few significant results were found regarding inter-syndrome differences. Thus, these features might not be significant enough to differentiate pathological subjects, and other features should be considered in future investigations. As shown in Table 6.9, it is noteworthy that differences in acoustical features still exist when only PA subjects were considered. Indeed the number of such features increases, showing more differences in F2 and F3 with respect to CS.

Tables 6.2 and 6.3 also confirm that entropy indexes (CI MSE and Max MSE) could discriminate between CS and pathological voices. In CS, both CI MSE and Max MSE values were found to be lower than in pathological cases, as shown in Tables 6.2 and 6.3. Moreover, for some numbers, the entropy indexes give significant results even when the acoustical features obtained with BioVoice did not highlight differences. This suggests that the entropy measures may add useful information to the considered acoustical features. However, as opposed to acoustical features, the entropy features do not have a direct and clear correlation with physiological mechanisms.

We remark that, to the best of our knowledge, the present work is one of the few concerning the use of quantitative acoustical analysis applied to genetic syndromes during a specific speech task. The analyses carried on are thus a first attempt to find acoustical differences among genetic syndromes other than the analysis of sustained vowels only. The results agree with the findings obtained with the analysis of sustained vowels [Hidalgo-De la Guía et al., 2021a, Moura et al., 2008, Deller et al., 2000], confirming that acoustical analysis could provide helpful information for the characterization

of speech phenotype for the considered syndromes. Moreover, the significant acoustical features may be used as an input for machine-learning models to improve their performance in the characterization and assessment of genetic syndromes [Frassineti et al., 2021e].

Finally, we point out that another advantage of quantitative acoustical analysis could be the longitudinal monitoring of the same subject in order to evaluate or detect any change in his/her vocal features over time. Moreover, the characterization of language phenotype in genetic syndromes could be helpful for the identification of subjects that need dedicated logopaedic or rehabilitation intervention and in follow-up procedures. Though innovative, the work presented here has some limits. First, the number of subjects is quite low, because SMS and NS are rare syndromes [Hidalgo-De la Guía, 2021a, Lazzaro et al., 2020, Hidalgo-De la Guía, 2021b]. The recruitment of a larger number of subjects could also allow intra-syndrome differences investigation to evaluate if acoustical differences could be highlighted according to the severity level of the syndrome. Furthermore, future work could be devoted to disentangle differences due to specific altered biomechanical properties of the vocal tract, cognitive delay, or both. To this aim, more clinical and psychological information might be added (e.g., skull dimension, anatomical properties of the vocal tracts, IQ scores, etc.) in a multidimensional analysis perspective.

Moreover, the work presented here concerns Italian speakers only, and most likely the relevant parameters would be different for other languages. We limited the analysis to a single task: the list of numbers from 1 to 10, as it is included in the SIFEL protocol. Maybe other tasks could be more appropriate to find significant differences. Another possible task could be the repetition of words and un-words that have already shown differences between normal speakers and pathological ones [Tressoldi et al., 2001]. Another limitation of this work may be the use of commercial smartphones for audio recordings rather than a professional microphone [Penney et al., 2021, Lebacq et al., 2017]. However, even with smartphones, the acoustical analysis was able to detect differences, thus allowing their possible use also in a non-controlled environment such as home monitoring [Manfredi et al., 2017].

6.4.1. Conclusions

Though with the above mentioned limitations, this work proposes a first proof of concept for the analysis of a speech task to improve speech phenotyping in genetic syndromes. A first attempt is presented concerning the usefulness of quantitative acoustical analysis for the characterization of speech phenotypes in genetic syndromes. Preliminary encouraging results confirm that acoustical measures could add helpful information for Down, Noonan, and Smith-Magenis syndromes. Being completely non-invasive, acoustical analysis might significantly contribute to the clinical assessment of such subjects, also after surgical, pharmacological, psychological and logopaedic treatments as well as for long-term monitoring of the acoustical quality of the voice of these subjects. Several specific differences were found between pathological and control cases, showing a peculiar picture for each syndrome considered. Though non-optimal, the use of smartphones may allow the analysis both in a controlled environment, such as specialized clinics and hospitals, and/or in-home monitoring, after careful training of the parents or tutors of the subject. Therefore, this study brings innovative elements that combine acoustical analysis and sample entropy techniques in the study of genetic syndromes, showing their potentiality and possible usefulness in clinical practice. This analysis opens the way to the use of acoustical features as input for artificial intelligence models devoted as diagnostic tool for an automatic characterization of the speech phenotype in such genetic syndromes. A first attempt regarding this possibility, will be briefly presented in the next section 6.5.

6.5 Automatic classification of vocal patterns as a diagnostic tool in patients with genetic syndromes

As already explained in Section 6.1, perceptual and acoustical analysis of voice could be helpful for the evaluation of specific voice characteristics as a non-invasive approach to the assessment of genetic syndromes. More than 240 genetic syndromes have distinctive abnormalities of voice quality, significant enough to be considered as diagnostic indicators [Hamosh et al., 2005]. For some genetic syndromes the existence of a specific language phenotype obtained by acoustical analysis was already discussed in Section 6.1, suggesting that acoustical analysis could be helpful for an early intervention in patients with speech impairments, to improve their communication skills and reduce speech deficits [Moura et al., 2008]. Based on the above mentioned evidences, some genetic abnormalities of a specific phenotype are expected to determine a specific vocal phenotype.

Therefore, vocal characterization could represent a useful tool in the diagnostic process and in defining the severity of some clinical pictures. To this aim, machine-learning methods and supervised classifiers are applied here to acoustical parameters estimated with two analysis tools: Praat and BioVoice [Boersma and Weenink, 2018, Morelli et al., 2021]. Being based on a non-invasive and easily administered tests, this approach could be helpful for obtaining additional features useful for diagnosis and for the automatic classification of different syndromes.

6.5.1 Material and Methods

Data were collected at the Università Cattolica del Sacro Cuore, (Roma), Faculty of Medicine and Surgery [Calà, 2022]. Machine-learning methods are applied to several acoustical parameters estimated from the vocal emissions of a set of 72 subjects (36 male and 36 female, age range 4-33 years, mean 14±7 years), affected by 5 different genetic syndromes. Specifically, the dataset consists of: 22 subjects with Down syndrome (DS); 17 with Noonan syndrome (NS); 19 with Costello Syndrome (CoS) [Myers et al., 2014]; 10 with Smith-Magenis syndrome (SMS) and 4 with Cornelia de Lange syndrome (CdLS) [Moore, 1970]. However, the CdLS syndrome was excluded from the analysis due to the small number of subjects in this class. The vocal samples come from a previous study based on the SIFEL protocol [Ricci and Maccarini, 2002]. After a training phase of the subject, the recorded audio files consist of the vowel /a/ sustained for at least 4 seconds. Recordings were obtained using a portable DAT (Digital Audio Tape) in a controlled environment (environmental noise < 40dB), with the microphone set at 15 centimetres from the subject's lips and with an angle of 45°. The sampling rate was 44100 Hz. Moreover, in the same sessions, the Italian word /aiuole/ (flower beds) as well as the vowels /i/, /u/ /o/ and /e/ were recorded. However, in this work the acoustical analysis with BioVoice was performed only on the vowel /a/, because for other vowels and words, some of the recordings were corrupted or no more available. For the other vowels and words, only the acoustical analysis previously performed by Praat [Boersma and Weenink, 2018] was available. The quasi-stationary central part of each sustained vowel (about 3s of duration) was manually extracted by an expert, disregarding onset and offset.

For the acoustical analysis and classification we considered here both the previously collected dataset of parameters estimated with Praat and new estimates obtained with the BioVoice tool [Morelli et al., 2021, Manfredi et al., 2015]. Only the sustained vowel /a/ was considered. With Praat, the following 34 acoustical parameters were taken into account: mean, standard error, coefficient of variation, maximum and minimum of the fundamental frequency F0; Jitter (local, absolute, Relative Average Perturbation, DDP and PPQ5, where PPQ is Period Perturbation Quotient); Shimmer (%, dB, APQ3,

APQ5, APQ11, DDA, where APQ is the Amplitude Perturbation Quotient); mean Noise to Harmonic Ratio (NHR); mean Harmonic to Noise Ratio (HNR); the first four formants (F1, F2, F3 and F4); four clinical features: gender, age, weight and body mass index.

With BioVoice we extracted 24 acoustical features. Analysis is performed distinguishing between infants (<12 years) and adults [Morelli et al., 2021] and in the case of adults between male and female. The 24 acoustical parameters from BioVoice are: maximum, minimum, mean, median and standard deviation for F0 and formants F1, F2 and F3; T0min and T0max for F0; jitter; Normalized Noise Energy (NNE). Four clinical features: gender, age, weight and body mass index (BMI) were also included. In a first step, we compared the acoustical parameters in common between BioVoice and Praat. Then, we used those parameters considering separately each syndrome subgroup. All features except gender (0=male, 1=female) were normalized to zero mean and unit variance and the corresponding feature matrix was applied as input to the following supervised classifiers: k-nearest neighbours (KNN), support vector machine (SVM) and ensemble methods (RUSBoost, AdaBoost and Random Forest) [Hastie et al., 2001]. These methods are implemented under MATLAB 2020b computing environment.

K-fold cross validation (k=5) and Bayesian Optimization were applied for the selection of the hyper-parameters of the models [Calà, 2022]. The optimization was performed considering the highest global Accuracy as validation metric (i.e. the average Accuracy between the four classes). To improve the classifier's performance the ReliefF algorithm was used as feature selection method [Robnik-Šikonja and Kononenko, 2003]. During the model selection process we also varied the number of input features for the classifiers. All the experiments were repeated 5 times, to take into account possible variations of the performance due to the random selection of the subjects during cross-validation.

We did not find significant differences in the performances (Accuracy <5%). Finally, we performed the same experiment on the Praat dataset, considering also features from the vowels /a/, /i/ and /u/. In this case the features given by the formant ratios between vowels were added (e.g., $F1_{[a]}/F1_{[u]}$) [Boersma and Weenink, 2018]. As said before, this analysis could not be performed with BioVoice due to missing data.

6.5.2 Results

Table 6.13 shows the comparison between Praat and BioVoice concerning the vowel /a/. We used a two-sample t-test with level of significance α=0.05. We checked the hypothesis of normality by Shapiro-Wilk Test (level of significance α=0.05). Table 6.14 shows the True Positive Rate (TPR) and the False Negative Rate (FNR) for the four genetic syndromes. With BioVoice the 10 features obtained for the best model were: $T0_{maxF0}$ /a/, gender, age, median F3 /a/, BMI, min F1 /a/, $T0_{minF0}$ /a/, min F0, jitter and weight. The best model for BioVoice was a KNN with a Global Accuracy of 53.1%. Instead with Praat the best model was made of 15 features: gender, mean F1 /a/, age, mean F2 /a/, BMI, max F0 /a/, min F0 /a/, weight, mean F0 /a/. median F0 /a/, Shimmer /a/ APQ11, Shimmer /a/ APQ5, Shimmer local /a/, mean F4 /a/, Shimmer /a/ DDA. The best model with Praat was a KNN with 52.9% of Global Accuracy. The features used after the selection process are listed in descending order according to their relevance.

Table 6.13. Vowel /a/ - Comparison between BioVoice and Praat on the 4 syndromes. Statistically significant differences are highlighted in bold. [Frassineti et al., 2021e].

	Syndrome (p-value)			
Feature	**DS**	**NS**	**CoS**	**SMS**
Median F0 /a/	0.91	0.74	0.99	0.77
Mean F0 /a/	0.80	0.80	0.95	0.66
Min F0 /a/	**0.01**	0.05	**p<0.01**	0.13
Max F0 /a/	**p<0.01**	0.44	**0.02**	0,16
Mean F1 /a/	0.55	0.43	0.92	0.56
Mean F2 /a/	**p<0.01**	**p<0.01**	**0.03**	0.11
Mean F3 /a/	**p<0.01**	0.12	0.23	**p<0.01**

Table 6.14. Vowel /a/ - Comparison between BioVoice and Praat - Results of k-fold cross validation. [Frassineti et al., 2021e].

Genetic Syndrome	**BioVoice**		**Praat**	
	TPR	**FNR**	**TPR**	**FNR**
DS	61.9%	38.1%	63.6%	36.4%
NS	26.7%	73.3%	17.6%	82.4%
CoS	68.4%	31.6%	73.7%	26.3%
SMS	55.6%	44.4%	40.0%	60.0%

Table 6.15 shows the results obtained for the four genetic syndromes considering all the available Praat features for vowels /a/, /u/ and /i/.

Table 6.15. Vowels /a/, /i/ and /u/ - KNN's Multiclass confusion matrix with Praat parameters. Main diagonal: TPR for each class. Other values: FNR for a single class. [Frassineti et al., 2021e, Calà, 2022].

	Predicted Class			
True Class	**DS**	**NS**	**CoS**	**SMS**
DS	68.2%	13.6%	18.2%	0%
NS	17.6%	64.7%	17.6%	0%
CoS	31.6%	5.3%	63.2%	0%
SMS	20.0%	10.0%	10.0%	60.0%

The best model was a KNN with Global accuracy 64.7% [Calà. 2022]. In this case, the following 15 features were selected: mean F1 /a/, age, gender, formant ratio $F1_{[a]}/F1_{[u]}$, max F0 /a/, mean F2 /a/, Shimmer APQ11 /a/, mean F0 /a/, median F0 /a/, min F0 /a/, Shimmer /a/ (dB), BMI, Shimmer APQ5 /a/, weight, Shimmer /a/ (local).

6.5.3 Discussion and Conclusions

This work presents preliminary results concerning the discrimination among some genetic syndromes: Down Syndrome, Noonan Syndrome, Costello Syndrome and Smith-Magenis Syndrome. The analysis was performed with acoustical parameters estimated on the sustained vowel /a/ with BioVoice and Praat and applying machine-learning models. The aim of this work was the definition of a proper language phenotype able to distinguish the genetic syndromes considered. The results

shown in Table 6.14 and 6.15 confirm a possible relationship between genetic syndromes and their specific acoustical characteristics. The results obtained with BioVoice and Praat are comparable. Statistical analysis highlights some differences between the two tools as far as the estimation of formants F2 and F3 for some syndromes is concerned (Table 6.13, p-values <0.05). This might be related to different techniques for formants estimation implemented in the two tools, as discussed in [Morelli et al., 2021]. Moreover, differences between BioVoice and Praat exist concerning F0 max and min. This could be due to different ranges for F0 estimation defined by the two software tools. We remark that with BioVoice the selection of the frequency range for adults (male or female), infants and newborns is automatically made by BioVoice, while Praat requires some skill of the user to manually set the best frequency range. However, the results shown in Table 6.14 are preliminary, suggesting that the analysis of the vowel /a/ alone might not be enough for defining a vocal phenotype (TPRs<50%). This is confirmed in Table 6.15, where the acoustical analysis of vowels /i/ and /u/ performed with Praat was added for all the syndromes, giving Accuracy>50%. In particular, the formant ratio $F1_{[a]}/F1_{[u]}$ was classified as one of the most relevant features by the ReliefF algorithm. This result suggests that a multi-vowel analysis might add more information than a single vowel analysis and should be preferred for the characterization of these genetic syndromes. Our results also confirm evidences previously found for some genetic syndromes. Indeed, for DS, NS and SMS acoustical analysis was already proved useful to find differences between pathological and control groups [Hidalgo-De la Guía et al., 2021a, Hidalgo-De la Guía et al., 2020, Lazzaro et al., 2020]. Table 6.15 also shows that SMS has the lowest false negative rate (0%), confirming that acoustical analysis can provide characteristics strictly related to the pathology [Hidalgo-De la Guía, 2020]. Our results suggest that acoustical analysis could be useful also for CoS. Indeed, as shown in Table 6.15, the false negative rates between CoS and NS were 5.3% and 17.6% respectively, thus acoustical analysis might be useful to discriminate between these two syndromes [Myers et al., 2014].

Our results are preliminary and further study is required to confirm them. First, the number of subjects was poor, thus more cases must be recruited especially for SMS and CdLS. Moreover, we did not perform a comparison between pathological subjects and control cases. This will be done in future work, also taking into account previous studies that already presented such differences for some genetic syndromes [Hidalgo-De la Guía et al., 2021a, Hidalgo-De la Guía et al., 2020, Lazzaro et al., 2020]. Considering the promising results obtained, further studies will be made to investigate if some of the acoustical features could be specific of a single genetic syndrome. The acoustical analysis of vowels /i/ and /u/ made with the Praat dataset was found useful, therefore we are planning to perform the same analysis with BioVoice on the same recordings, when available, and/or new ones. Another limit of the work presented here is the wide age range of the subjects, also due to the low number of cases in some syndromes (e.g. CdLS or SMS). If other subjects will be available, a more detailed analysis at different age ranges will be made. If successful, acoustical analysis may be included in the process of differential diagnosis as a completely non-invasive approach to detect specific acoustical characteristics related to speech or phonation impairment for several genetic syndromes, along with e.g. the analysis of facial characteristics and expressions [Bandini et al., 2016].

The work presented in this Section is a first step towards the analysis of the complex mosaics behind the detection of “voice” phenotypes related to some genetic syndromes. Preliminary results suggest that acoustical parameters and supervised classifiers might provide additional information about genetic syndromes through the characterization of voice. Future work will be devoted to the definition of a protocol for data recording and will concern a larger number of subjects and syndromes, as well as different supervised classifiers and feature selection approaches.

7. Neonatal Sepsis and Neurodevelopment: forecasting BAYLEY-III scores through EEG- and HRV-based regression analysis.

Some contents in this chapter are based on the following Master's Degree Thesis:

- *Parente, Angela. Thesis title: Machine learning techniques for studying the influence of sepsis on neurodevelopment in preterm infants: application to the Neonatal Intensive Care Unit AOU Careggi, Florence. Università degli Studi di Firenze (Firenze, Italy), Scuola di Ingegneria, Curricula Biomedical Engineering, date of discussion: 07/21/2022. Supervisors: Prof. Claudia Manfredi, Prof. Antonio Lanatà. Co-Supervisor: Lorenzo Frassineti. [Parente, 2022].*

In this chapter another application of Artificial Intelligence (AI) models on paediatric subjects is presented. The use of regression models is exploited to predict the neurodevelopmental scores of preterm newborns with sepsis and without sepsis. In adults, sepsis is defined as a life-threatening organ dysfunction caused by a dysregulated response to infection [Singer et al., 2016]. Although for newborns there is still no consensus definition of neonatal sepsis [Molloy et al., 2020], it can be defined as a diagnosis made in infants less than 28 days of life and consists of a clinical syndrome that may include systemic signs of infection, circulatory shock, and multisystem organ failure [Ershad et al., 2019]. The BAYLEY-III test was used to compute the scores in three different domains: cognitive, language and motor. The quantitative analysis was performed on EEG and ECG recordings acquired when the preterm infants reached the term age (i.e. > 37 weeks). This application is one of the first attempt to use regression models as a support tool for the neonatologists and the paediatric neurologists in the neurodevelopmental assessment of the preterm newborn with sepsis.

The chapter is organized as follows: an overview on neonatal sepsis and methods to detect early the sepsis episode or predict its effect on neurodevelopment is provided. Then, the framework used to develop the regression models is presented. Finally, the results and discussion about the use of AI regression models as support for the neurodevelopment assessment on newborn with sepsis are reported.

7.1. Introduction

Each year, almost 15 million of infants born premature, that is before the 37th week of pregnancy. They are about the 10% of the worldwide neonatal population [González et al., 2011]. At least 33% of hospitalizations in Neonatal Intensive Care Units (NICUs) are related to preterm newborns. Moreover, the preterm birth rate, defined as the ratio between preterm births and number of newborns born alive, has increased from the 9.6% in 2005 [Beck et al. 2010], up to 11.1% in 2010 [Blencowe et al., 2012].

Specifically, depending on the gestational age (GA) [World Health Organization, 2012], newborns can be classified as follows:

- At term: if the birth is between the 37th and 42nd gestational week
- Late Preterm: if the birth is between the 32nd and the 37th gestational week
- Very Preterm: if the birth is between the 28th and the 32nd gestational week
- Extremely Preterm: if the birth is before the 28th gestational week

It is important to remark that the risk of the death increases in newborns with a GA lower than 34 weeks [Marlow et al. 2005]. Moreover, neonatal deaths mainly occur during the first week of life (almost the 80%, [Lawn et al., 2014]). Another recognized classification is defined according to the birth weight [World Health Organization, 2012]:

- Low Birth Weight (LBW): newborns with a weight between 1501g and 2500g
- Very Low Birth Weight (VLBW): newborns with a weight between 1001g and 1500g
- Extremely Low Birth Weight (ELBW): newborn with a weight lower than 1000g

This classification is used in combination with the corresponding GA.

The survival rate of preterm newborns varies worldwide: in high-resources country such as Italy, the percentage of survived newborns after the first week, with a GA lower than 28 weeks, is ~~the~~ 90% [World Health Organization, 2012]. Instead in low-resources countries this percentage drops to 10%.

The death causes for newborns are several and no exhaustive worldwide list is provided. The most common ones are shown in Figure 7.1. A preterm birth may carry several complications such as: Necrotizing enterocolitis (NEC), Retinopathy of prematurity (ROP), Bronchopulmonary Dysplasia (BPD), that they are the most common causes of neonatal death, with almost 3 million of deaths per year, the second cause of death under 5 years of life [Blencowe et al., 2012].

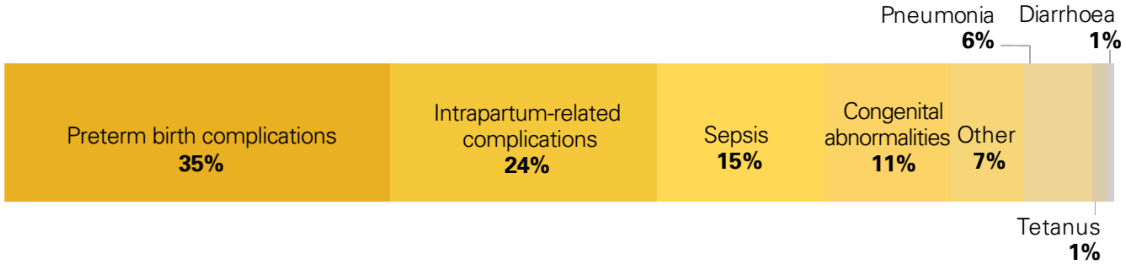


Figure 7.1. Global distribution of newborn deaths by cause (2018) [Unicef Report 2019].

As shown in Figure 7.1 among the death causes Sepsis represents the third most common cause. Infections during hospitalization represent a significant issue and challenge for healthcare and clinical staff. The incidence of infections in Europe is between 7% and 19%, in Italy it has been reported an incidence between 5-8%. The clinical signs of a neonatal infection are numerous and unspecific, and difficult to be detected. At the systematic level the newborn may show hypotension, hyperthermia or hypothermia, respiratory stress with apnoea, feeding difficulties, as well as an increasing demand of oxygen or assisted respiratory support [Bollani et al., 2016].

In the 45-55% a neonatal infection is due to sepsis, followed by the lower respiratory tract infections (16-33%), cutaneous infections (26.3%), urinary infections (8-19%) and meningitis (9.6%) [Bollani et al., 2016]. It has been estimated that the worldwide number of neonatal sepsis episodes are at least 1.2 millions each year [Weiss et al., 2020].

The mortality rate of neonatal sepsis varies between 4% and 50% [Balamuth et al., 2014, Ames et al., 2018, Prout et al., 2018]. Most infants that die for sepsis suffer from multiorgan disfunctions and refractory shock, and most of the deaths occur within the first 48-72 hours [Weiss et al., 2017, Morin et al., 2016]. It is noteworthy that with an early diagnosis and treatment, rarely a preterm newborn shows long-term effects due to the infection. However, it is still not easy to detect and evaluate any early symptoms of the infections, especially from the clinical signs only.

Usually, neonatal sepsis is divided into two categories:

- **Early Onset Sepsis** (EOS), where symptoms occur until the first 72 hours of life (85% until the first 24 hours). Usually, the infection is transmitted from the mother to the newborn. The death rate is higher for sepsis caused by Gram-negative microorganisms (e.g., Escherichia coli and Haemophilus influenzae) [Bollani et al., 2016]. Several risk factors contribute to the development of EOS: low APGAR score, maternal urinary infections, prematurity, asphyxia etc. [Bollani et al., 2016].
- **Late Onset Sepsis** (LOS), that occurs after the first 48-72 hours of life, mainly caused by external microorganisms. For VLBW newborns, it is also possible to define the very-late-onset Sepsis (VLOS), after the first 60 days of life and late-late-onset sepsis (LLOS), where sepsis occurs three months after birth. The incidence of LOS increases drastically in NICUs. In fact, among the risk factors it is possible to identify: mechanical ventilation, antibiotic therapy prolonged over time, cardiac deficits and any surgical intervention [Bollani et al., 2016].

The blood culture test is the gold standard for the diagnosis of sepsis. The exam allows to recognize the aetiology, thus defining the corresponding treatment. However, often blood culture may produce false negative and positive results, thus it is not always clear when the antibiotic treatment should start [Bollani et al., 2016]. Moreover, the blood culture test requires time, thus the clinical staff might start the antibiotic therapy before the result of the test.

To support the clinical staff in the detection or prediction of newborn at risk of sepsis, in the last years several AI techniques have been proposed. To develop such methods three different datasets have been considered in the literature:

- The Medical Information Mart for Intensive Care III (MIMIC-III database), where almost 8000 data regarding newborns are included [Johnson et al., 2016].
- The dataset provided by Lopez-Martinez et al. [López-Martínez et al., 2019], where 555 newborns were included of which 34% with neonatal sepsis.
- The dataset provided by Masino et al. [Masino et al., 2019], where 1100 newborns without sepsis and 375 with sepsis are included. The newborns with sepsis were further divided into two categories: "*with a blood culture positive*" (110) and "*clinically positive*" (265). Clinically positive cases are those with a culture negative but with the antibiotic therapy started at least 120 hours before. They collected 36 clinical features spanning different domains (e.g., laboratory features such as creatinine or glucose, as well as vital signs features such as diastolic blood pressure or heart rate etc.).

The aim of the methods developed in this field is mainly the discrimination between newborns with sepsis and without sepsis. They can be developed and validated in a similar way as the seizure detectors discussed in the previous chapters of this PhD thesis (from Chapter 1 to 5), considering the sepsis detection or prediction as a supervised classification problem. The most common metrics used for the evaluation of sepsis detectors are: the Positive Predictive Value (PPV) and the Negative Predictive Value (NPV), as well as Accuracy, Sensitivity and Specificity. However, the most used metric is the AUROC (or AUC). More details regarding the metrics considered in literature are reported in Chapter 1 and 2 of this PhD thesis.

A comprehensive discussion about the main works on this topic is out of the aim of this PhD thesis, thus only the most relevant are reported here. Furthermore, only the works who used the three datasets described in this section will be introduced. We point out that there are less research works concerning

neonatal sepsis detection and prediction than for adults. Tahkur et al. in two different works [Tahkur et al., 2017, Tahkur et al. 2020] used the MIMIC-III database to develop and validate Machine-Learning (ML) models to predict neonatal sepsis. In [Tahkur et al., 2017], they obtained an AUC of 76% evaluating 1446 newborns, 179 with sepsis, by a Logistic Regressor model and the following features: birth weight, heart rate, body temperature, oxygen saturation and blood pression. Instead in [Tahkur et al., 2020], they showed that using only the body temperature as feature, it was still possible to obtain promising performances in terms of Accuracy. Song et al. [Song et al., 2020], evaluated the performance of ML models to detect LONS newborns using the MIMIC-III data. They obtained an AUC of 86% using a Logistic Regressor model.

Regarding the dataset proposed by Lopez-Martinez et al. [López-Martínez et al., 2019], the same authors obtained an AUC of 92%, with an Artificial Neural Network using 27 clinical features. Masino et al. [Masino et al., 2019], on their dataset, obtained an average AUC of 80-82% for "*clinically positive*" cases and 85-87% when they considered also the "*culture positive*" cases. All these results confirming that AI models may be used in the future as a support for an early detection of newborns at risk of sepsis.

In the last years, the improvements in the management and treatment, as well as the survival procedures for preterms, allowed a higher percentage of survived preterm newborns with sepsis. However, it has been found that preterm newborns are more sensible to neurodevelopmental delay or diseases, when compared to at term newborns. In fact, it has been noticed that more than 25% of newborns with a GA between the 28^{th} and 32^{nd} week show a delay on neurodevelopment, usually linked to several degrees of impairment [Lawn et al., 2014]. Recently, it has been argued that, in the survived newborns, sepsis may have a negative impact on their neurodevelopment. In fact, sepsis may cause significant alterations to cerebral networking in the neonatal period and could be harmful for brain development [Mukhopadhyay et al., 2020, Alshaikh et al., 2013, Ortgies et al., 2021, Pek et al., 2020].

Furthermore, the early detection of neurodevelopmental disorders or delay is of utmost importance in the clinical practice, as the first two years of life are considered the most vulnerable and critical period for the neurodevelopment [Scher, 2021]. Thus, the newborn at risk or with sepsis-related damages should be identified as soon as possible in order to define the best neuro-rehabilitative program [Scher, 2021].

To monitor the neurodevelopment and detect any abnormal behaviour, the clinical staff often makes use of neurodevelopmental scales such as the Bayley Scales of Infant and Toddler Development or Griffiths Mental Development Scales [Del Rosario et al., 2020]. In general, these scales consist of a list of tests and tasks that the physician administer or verify on the infants at different follow-up periods, usually from the 3-6 months after birth up to 18-24 months after birth (considering the corrected age for preterm). As an example, the BAYLEY-III scale provides assessment on three different subscales of the neurodevelopment: cognitive, language and motor [Del Rosario et al., 2020]. Moreover, for BAYLEY-III two more subscales indirectly administered are usually considered: the social-emotional and the adaptive behaviour scales.

Each scale comprises a different number of items~~,~~ and their administration is flexible but with a standard order. Moreover, the number of items varies according to the age of the subject. For example, the cognitive scale (maximum number of items 91) evaluates elements such as the development of spatial exploration, memory, manipulation, relationship between objects and concepts, information comprehension. This is usually the first scale administered during the evaluation since it requires a high effort for the subject.

According to the BAYLEY-III scores, it is possible to quantify the level of impairment of the subject. Usually scores below 85 and 70/75 denote a moderate or severe impairment, respectively [Del Rosario et al., 2020], while scores between 85/90-100 and higher are associated with normal conditions.

Recently, some works in the literature proposed Artificial Intelligence (AI) methods to predict the neurodevelopment scores, using mainly Electroencephalographic (EEG) signals acquired from the newborns during or immediately after their stay in NICU [Alotaibi et al., 2022]. Thus, these methods could be used as a support for the clinical staff in the early detection of newborns at risk of neurodevelopmental disorders. Moreover, it is well known that the neurodevelopment itself is altered by preterm birth [Yiallourou et al., 2013] and changes could be detected by the analysis of the Autonomic Nervous System (ANS) [Thiriez et al., 2015]. Also Heart Rate Variability (HRV) analysis reflects ANS activity, thus it can provide information about its development [Ask et al., 2019]. HRV analysis can be obtained by Electrocardiographic (ECG) signals that usually are easily obtainable in the clinical practice, being less invasive and cheaper than EEG. However, to date, only few works investigated HRV features and AI methods as predictors of neurodevelopmental scores [Ask et al., 2019]. Moreover, to the best of our knowledge, the work presented in this PhD thesis is the first one concerning newborns with sepsis. Thus, here we evaluated if EEG or HRV features as input features of regression models may provide a reliable estimation of BAYLEY-III scores for preterm newborns with and without sepsis obtained during follow-ups at 6- and 12- months.

This chapter is organized as follows: Section 7.2 describes the dataset, the pre-processing applied to EEG and ECG signals, the features extracted, and the validation scheme adopted on regression models to evaluate their performance in predicting BAYLEY-III scores. In Section 7.3, results are shown. Section 7.4 is devoted to the discussion and conclusions about the use of EEG and HRV as predictors of neurodevelopmental scores.

7.2. Material and Methods

A dataset of EEG and ECG recordings was collected at the Neuro-physiopathology and Neonatology Clinical Units of AOU Careggi (Firenze, Italy). EEG and ECG were synchronously recorded using Nemus ICU Galileo NT Line system (EB Neuro S.p.A., Firenze, Italy). The dataset was collected between 2018 and 2022. The length of EEG and ECG signals was about 54±9 minutes, with a sampling frequency of 128Hz. The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of Careggi University Hospital, Firenze, Italy. It consists of 64 preterm newborns with gestational age (GA) between 24 and 31 weeks (27.8 ± 1.8 weeks). The age of newborns was between 37 and 43 weeks (38.5 ± 1.5 weeks). Thus, all the newborns were analyzed when they reached the corrected term age (> 37 weeks). Regarding the ECG signals only 48 of 64 subjects were considered, since for the others the ECG signals were corrupted by noise. For the EEG recordings, 38 out of 64 subjects had sepsis during hospitalization, while for the ECG recordings, 27 out of 48 patients had sepsis during hospitalization. The distribution of cases as far as age at time of recording, GA, and sepsis and no-sepsis, are shown in Figure 7.2. Regarding the definition of sepsis used here, both EOS and LOS were included, referring to them with the same notation “sepsis”, without considering how it was diagnosed (e.g. with positive blood culture or by clinical evaluation).

The BAYLEY-III scales were administered by an expert psychologist of the AOU Careggi staff. Both the cognitive, language and motor scores were collected at 6-months and 12-months follow-up. had

Here, as in [Del Rosario, et al., 2020], a moderate/severe impairment refers to the subjects with BAYLEY scores lower than 85. Further details about the subjects involved in the study are reported in Table 7.1 for the EEG recordings and in Table 7.2 for the ECG recordings. In Table 7.3 and 7.4 the BAYLEY-III scores are shown for the EEG and ECG recordings, respectively. Pearson's $\chi 2$ test and Mann-Whitney tests confirmed no statistical differences as far as gender and age are concerned between the two groups considered: sepsis and no-sepsis (Mann-Whitney test, level of confidence 0.05, both for EEG and ECG cohorts).

Table 7.1. EEG dataset details, μ=mean, σ= standard deviation.

Group	GA μ±σ weeks	Age at EEG μ±σ weeks	Cases (M/F)
Sepsis	27.3 ± 1.8	38.8 ± 1.2	(25/13)
No-sepsis	27.8 ± 1.4	39.0 ± 1.8	(11/15)

Table 7.2. ECG dataset details, μ=mean, σ= standard deviation.

Group	GA μ±σ weeks	Age at ECG μ±σ weeks	Cases (M/F)
Sepsis	27.5 ± 2.0	38.7 ± 1.3	27 (17/10)
No-sepsis	28.1 ± 1.5	38.8 ± 1.7	21 (8/13)

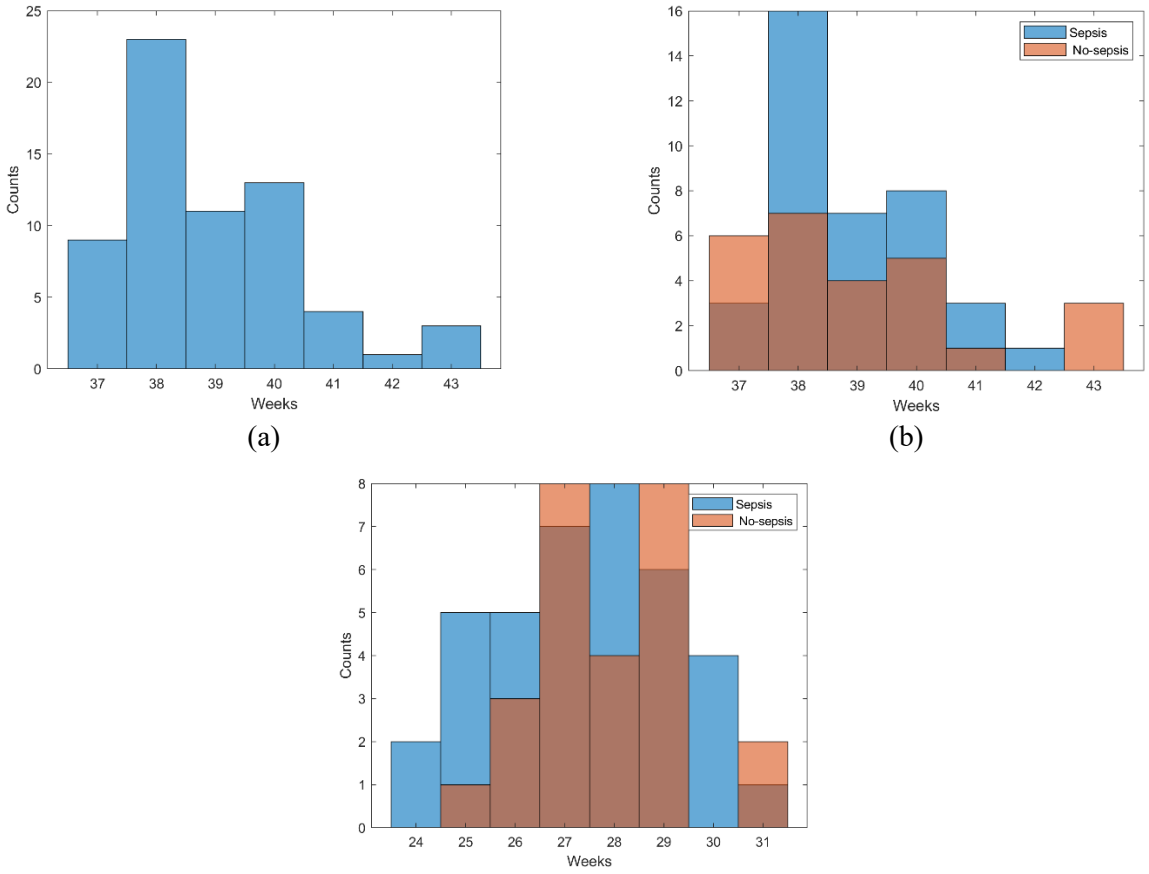


Figure 7.2. Histogram distribution of the EEG cohort. (a) The age in weeks at the time of EEG recordings and (b) the age at the time of the EEG recordings, for Sepsis and No-sepsis cases. (c) the GA distribution for the EEG cohort, for Sepsis and No-sepsis cases. [Parente, 2022].

As this study is still ongoing, only the scores for 6- and 12- months were available for a large enough number of infants. For the ECG cohort, for 6-months scores about all the Bayley-III evaluations were available for each domain, and for 12-months just 38 evaluations were available. For the 18-24-months scales, data and results will be presented in future developments of the research.

As shown in Table 7.3, for the EEG cohort we found statistical difference in terms of BAYLEY-III scores between sepsis and no-sepsis groups for the cognitive scale at 6- and 12-months and for the language scale at 12-months, respectively (Mann-Whitney test, level of significance 0.05). The number of cases varies slightly among the scales, because for some subjects it was not possible to administer all the tests. Thus, for the 6-months cognitive scale 62 of 64 subjects were considered. Instead, as shown in Table 7.4, for the ECG cohort, we found statistically significative difference in terms of BAYLEY-III scores between the two groups for the cognitive scale and for the language scale at 6- and 12- months respectively (Mann-Whitney test, level of significance 0.05).

Table 7.3. BAYLEY-III scores for all the sub-scales for both groups in EEG cohort. iqr is the interquartile range. The p-value is related to the Mann-Whitney test performed. m=months. (*) denotes a significant p-value (level of significance 0.05). [Parente, 2022].

	Group Median (iqr)			
BAYLEY-III scores	**Sepsis**	**No-sepsis**	**p-value**	**Cases (with Sepsis)**
Cognitive 6m	90 (15)	100 (10)	0.006*	62 (36)
Language 6m	77 (11)	83 (13)	0.07	61 (35)
Motor 6m	87 (13)	92 (9)	0.06	62 (36)
Cognitive 12m	90 (15)	100 (15)	0.04*	52 (31)
Language 12m	86 (12)	89 (8)	0.004*	52 (31)
Motor 12m	79 (9)	85 (12)	0.26	52 (31)

Table 7.4. BAYLEY-III scores for all the sub-scales for both groups in ECG cohort. iqr is the interquartile range. The p-value is related to the Mann-Whitney test performed. m=months. (*) denotes a significant p-value (level of significance 0.05).

	Group Median (iqr)			
BAYLEY-III scores	**Sepsis**	**No-sepsis**	**p-value**	**Cases (with Sepsis)**
Cognitive 6m	90 (15)	100 (11)	0.01*	48 (27)
Language 6m	77 (12)	83 (11)	0.06	48 (27)
Motor 6m	87 (12)	91 (12)	0.14	48 (27)
Cognitive 12m	93 (20)	102 (15)	0.15	38 (27)
Language 12m	86 (10)	90 (8)	0.01*	38 (27)
Motor 12m	79 (9)	86 (10)	0.29	38 (27)

7.2.1 EEG analysis

With the double banana montage for the EEG recordings, the following derivations were considered: Fp2-C4, C4-O2, Fp2-T4, T4-O2, Fp1-C3, C3-O1, Fp1-T3, T3-O1 [Parente, 2022]. To enhance the Signal-to-Noise ratio, the signals were filtered with a pass-band FIR filter of order 2000 in the range 1-45Hz. Because there is no unanimous consensus on the optimal epoch length to characterize neonatal EEG for neurodevelopment assessment, the following lengths have been evaluated: 4s, 8s,

16s, 128s and 256s [Lavanga et al., 2018, De Wel et al., 2017, Lavanga et al., 2017]. No overlap was applied to epochs. For each extracted epoch and for each derivation, the following single-channel EEG features were computed:

- Time-domain features: amplitude total power, amplitude standard deviation, skewness, kurtosis, mean of envelope, standard deviation of envelope. All the features were computed with the tool developed by O'Toole et al. [Toole and Boylan 2017].
- Frequency-domain features: spectral power, spectral relative power, spectral entropy, spectral difference, spectral edge frequency. [Toole and Boylan 2017].
- Entropy-domain features (Chapter 3 of this PhD thesis): multiscale sample entropy (MSE) analysis. For MSE the number of scales was set to 20, using as embedding dimension m=3 and tolerance threshold r=0.2. [Frassineti et al., 2021d, Costa et al., 2005]. For the multiscale procedure, the coarse-grained one was used [Frassineti et al., 2021d]. According to [De Wel et al., 2017], from MSE the Complexity Index, the maximum MSE among scales, the average slope for small scales (from 1 to 5) and for large scales (from 6 to 20) were computed and used as EEG entropy features. Thus, the original values of each MSE scale were not used in this analysis.

Furthermore, several multi-channel EEG brain network dynamics features were added [Frassineti et al. 2021a] as they showed promising results for the neurodevelopment characterization [Lavanga et al., 2018]. Thus, the following multichannel EEG features were computed:

- Mean and standard deviation of hemisphere coherence [Meijer et al., 2014].
- Connectivity features: mean connectivity coherence, maximum connectivity coherence, maximum frequency. These features were computed using the tool provided by O'Toole et al. [Toole and Boylan 2017].
- The Activation Synchrony index (ASI) and the Brain Synchrony Index (BSI) described in [Räsänen et al., 2013] and in [van Putten, 2007], respectively. Two EEG derivations were used to compute ASI and BSI indexes (Fp1-C3 and Fp2-C4, or C3-O1 and C4-O2). ASI and BSI were computed using the tool provided by O'Toole et al. [Toole and Boylan 2017].
- The Circular Omega Complexity (COC) [Baboukani et al., 2019] and the Hyper-Torus Synchrony (HTS) [Al-Khassaweneh et al., 2016], already introduced in Chapter 5 of this PhD Thesis.
- Multivariate Permutation Entropy (MVPE) [He et al., 2016], using as embedding dimension m=3.

Each single- and multi-channel feature was computed for the following frequency bands: Delta (0.5-4 Hz), Theta (4-7 Hz), Alpha (7-13 Hz), Beta (13-30 Hz). Overall, 467 EEG features were considered: 392 single-channel and 75 multi-channel. After the extraction of EEG features, the following statistics descriptors were applied for each subject: mean, median, standard deviation (std), kurtosis, skewness and interquartile range (iqr) [Frassineti et al., 2021b]. Thus, each EEG feature provided 6 different statistics descriptors. These features were considered as input to the regression models described in Section 7.2.3.

7.2.2 HRV analysis

To increase the Signal-to-Noise Ratio (SNR), before computing the Heart Rate Variability (HRV) features ECGs were pre-processed and filtered with a band-pass FIR filter in the bandwidth 0.05Hz-

45Hz. For HRV analysis a sliding time window of 5 minutes of duration without overlap was applied. Then inter-beat-interval (IBI) time series were obtained using the Pan-Tompkins'algorithm [Pan and Tompkins, 1985].

According to [Frassineti et al., 2021b, Olmi et al., 2022a], 82 HRV features were extracted in order to characterize the newborn's ANS. Specifically, the following features were considered:

- Time-domain features: Heart Rate (HR), standard deviation of normal-to-normal intervals (SDNN) and of RR intervals (SDRR), percentage of successive RR intervals > 50ms (pNN50), root mean square of successive RR interval differences (RMSSD), HRV triangular index (TRI), Triangular Interpolation of the NN interval histogram (TINN), Poincarè plot standard deviation along the line of identity (SD2), SD1/SD2, where SD1 is the Poincarè standard deviation perpendicular of the line identity (SD1SD2ratio), correlation dimension (CD) [Shaffer and Ginsberg, 2017].
- Frequency features: absolute power of Very Low (VLF), Low (LF) and High Frequencies (HF), the relative power for LF and HF (pLF and pHF), the total power (TP) and the LF HF ratio (LF/HF) [Shaffer and Ginsberg, 2017].
- Entropy-domain features: approximate entropy (ApEn), multiscale sample entropy from scale 1 to scale 20 (MSE1…MSE20), multiscale distribution entropy from scale 1 to scale 20 (MDE1…MDE20), multiscale fuzzy entropy from scale 1 to scale 20 (MFE1…MFE20) and Bubble Entropy [Manis et al., 2021]. The Complexity Index (CI) [Costa et al., 2005] for MSE, MDE and MFE was also computed.

As in [Statello et al., 2019, Lucchini et al., 2016], the LF range was adapted to (0.04-0.3) Hz, as well as the HR range (0.03-1.3) Hz. For multiscale entropy the coarse-grained procedure was applied [Frassineti et al., 2021d]. For all the entropy indexes, the embedding dimension was set to m=2. For MSE, a threshold value r=0.2 of the standard deviation of the epoch was considered. For MDE, we choose the number of bins B=512. For MFE the exponent n was set to 2 [Frassineti et al., 2021d].

As for the EEG features, after the extraction of HRV features, the following statistics descriptors were applied to each subject: mean, median, standard deviation (std), kurtosis, skewness and interquartile range (iqr) [Frassineti et al., 2021b]. Thus, each HRV feature gave rise to 6 different statistical descriptors. A matrix of size Nx492 was obtained, where N are the patients (48 at 6-months and 38 at 12-months) and 492 the overall statistics descriptors extracted from the original HRV features. These features will be used as input of the HRV-based regression models described in section 7.2.3.

7.2.3 Regression Analysis

As shown in Figure 7.3, before the regression model's validation, the training features were normalized (zero mean and unit variances). Moreover, the validation sets were transformed according to the training statistics in order to avoid any data leakage. Then, the highly correlated predictors were removed applying a threshold equal to |0.80| to the Pearson correlation coefficient of all the HRV predictors. The remaining features were ranked using the F-tests for regression (FSR) [Avots et al., 2022] or the ReliefF algorithm [Robnik-Šikonja and Kononenko,2003]. FSR tests the hypothesis that the response values grouped by the variable predictor values are drawn from populations with the same mean against the alternative hypothesis that the population means are different. Thus, the reordered HRV features were used as input predictors of support vector regression (SVR) models (linear and gaussian) and Ensemble regression models (Bag, LsBoost) [Hastie et al., 2001], starting

with models that considered only the first feature ranked by F-tests or ReliefF as far as using all the features.

To evaluate the developed models, the Leave-One-Subject-Out (LOSO) validation was performed using the Bayesian Hyperparameter Optimization, searching for the model with the lowest Mean Absolute Error (MAE, equation (7.1)). The number of iterations for Bayesian Optimization was set up to 500. Furthermore, the MAE metric was divided into MAE Sepsis (hereinafter MAE1, the MAE of regression models only on subject with sepsis) and MAE No-Sepsis (hereinafter MAE0).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{7.1}$$

Where n is the number of observations evaluated, y_i the i-th observed values and $\hat{y}_i$ its predicted values by the regression model. In our case the predicted values were the BAYLEY-III scores for each scale. During the LOSO validation the following hyperparameters were optimized:

- For the SVR models: box constraint and kernel scale [Olmi et al., 2022b]. Both Box constraint and Kernel scale were evaluated during the Bayesian Optimization in the range 10^{-5}:10^{5}.
- For the Ensemble models: number of learning cycles, minimum leaf size, maximum number of splits and learn rate (only for LsBoost models). During the Bayesian Optimization, the learn rate was evaluated in the range 0.001:1, the number of learning cycles between 10 and 300, the minimum leaf size between 1 and 50, while the maximum number of splits between 1 and the size of the current training set (e.g., for the cognitive scales at 6-months, the upper limit was set to 61).
- For both models (SVR and Ensemble): number of features, number of neighbours of ReliefF [Robnik-Šikonja and Kononenko, 2003]. During the Bayesian Optimization, the number of neighbours of ReliefF was evaluated in the range 1:15.

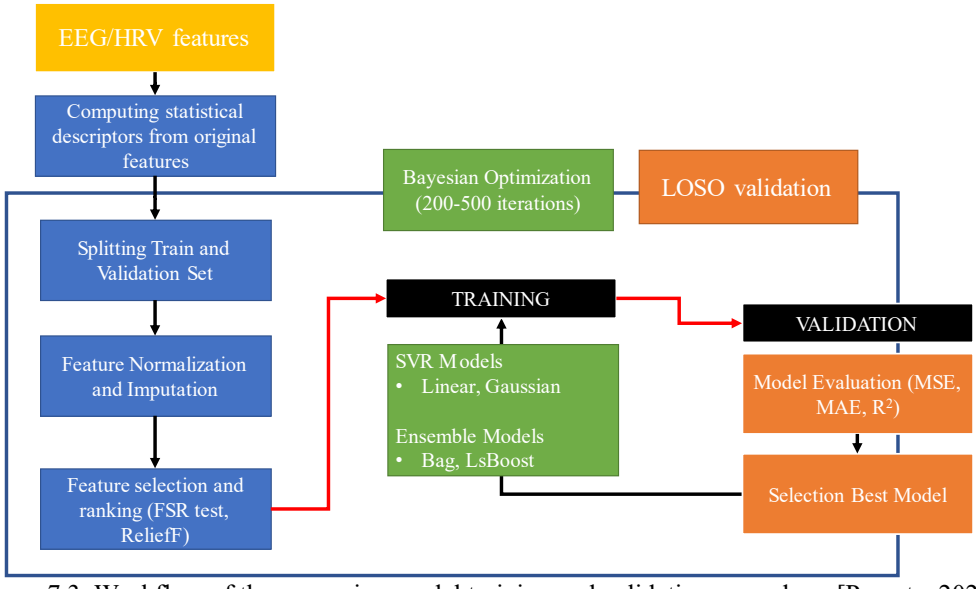


Figure 7.3. Workflow of the regression model training and validation procedure. [Parente, 2022].

To have a complete overview about the performance of the models, also the Mean Squared Error (MSE, equation (7.2)) for both groups and the R^2 parameter, was computed.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (7.2)$$

R^2, also known as the coefficient of determination, is a measure of the goodness of fit of a regression model. R^2 formula is reported in equation (7.3).

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (7.3)$$

Where SS_{res} is the residual sum of squares and SS_{tot} is the total sum of squares. An ideal SVR model should obtain a R^2 equals to 1.

This validation scheme was repeated for all the BAYLEY-III scales, both on EEG and ECG cohorts, generating EEG- and HRV-based regression models. The results related to regression analysis will be shown in Section 7.3.

7.3. Results

In this section the main results regarding the regression analysis described in section 7.2.3 are reported. The results of EEG-based regression models are related to 16s epochs that gave the best regression performance~~,~~ in terms of MAE, MSE and R^2. In Table 7.5 the results of the LOSO validation for all the BAYLEY-III scales, using EEG features, are reported. For the two follow-up periods considered (6- and 12-months), it is possible to identify the best results in terms of the lowest MAE. At 6-months follow-up, a MAE of 4.5 was obtained for the language scale, using a Bag regressor (Ensemble Models), FSR as feature selection method and with the following hyperparameters:

- number of learning cycles=63; minimum leaf size=2; maximum number of splits=19 and number of features=4.

Specifically, the four features used in the final model are: amplitude kurtosis T3-O1 (band β, iqr); envelope amplitude mean Fp2-C4 (band β, mean); amplitude skewness C4-O2 (band θ, std); amplitude skewness Fp2-C4 (band α, iqr).

For the 12-months follow-up the best results were obtained on the Language scale (Table 7.5, MAE=5.2). Such performance was achieved by a LsBoost regressor (Ensemble Models), ReliefF as feature selection method and with the following hyperparameters:

- learn rate=0.5870; number of learning cycles=104; minimum leaf size=13; maximum number of splits=1; ReliefF's k=5 and number of features=64.

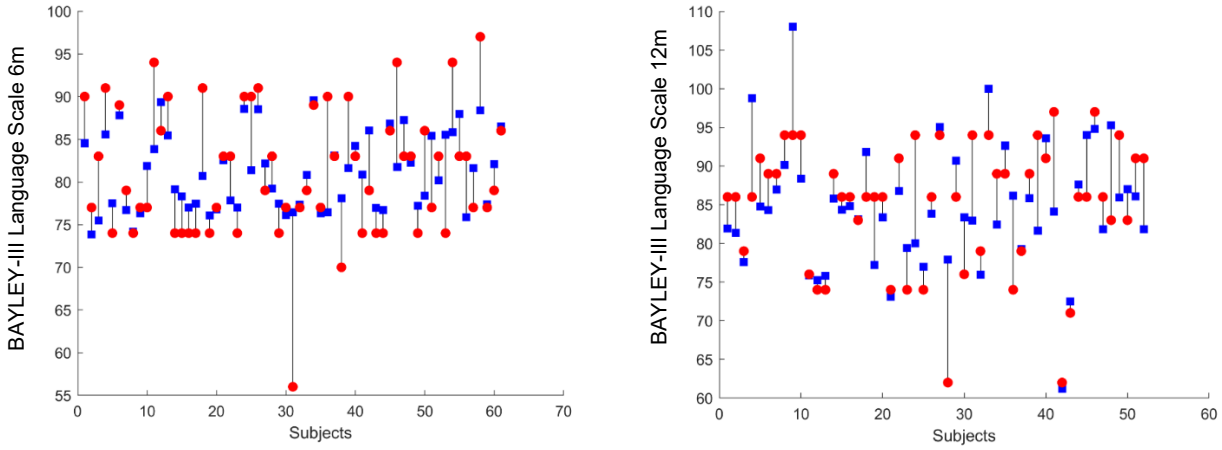
In Figure 7.4 the predictions of the two best EEG-based models in LOSO validation are shown. For the language scores at 6-months, the predicted scores (□, blue rectangles), the true scores (○, red circles) and their respective absolute error lines (black vertical lines) are reported in Figure 7.4a. The same is shown in Figure 7.4b for the language scores at 12-months.

The 12-months model was found more complex than to the 6-months one. In fact, it used 64 EEG features instead of the 4 features used for the other follow-up. Here, all the feature's names are not reported, however all of them belong to the single-channel domain. In particular, the first seven features are related to frequency-domain. Other models with less features than the best one were

investigated. However, all the models reached comparable performance using more than 60 features. These results confirm that the prediction of BAYLEY-III scales at 12 months by EEG features, obtained from recordings at term, remains a challenging and hardly interpretable task.

Table 7.5. Results of the LOSO validation for all the BAYLEY scales considered, using EEG features (m=months).

	Regressor metrics		
BAYLEY scales	**MSE**	**MAE (MAE0-MAE1)**	**R^2**
Cognitive 6m	53	5.8 (5.9 – 5.7)	0.14
Language 6m	35	4.5 (4.8 – 4.3)	0.36
Motor 6m	72	6.8 (5.6 – 7.9)	0.13
Cognitive 12m	92	7.6 (8.5 – 7.2)	0.10
Language 12m	45	5.2 (6.6 – 4.3)	0.33
Motor 12m	43	5.4 (5.7 – 5.2)	0.12



(a) (b)

Figure 7.4. (a) the predicted scores (□, blue rectangles) and the true scores (○, red circles) for the BAYLEY-III language scores at 6-months using EEG features. (b) predicted and true scores for the BAYLEY-III language scores at 12-months. Vertical black lines represent the absolute error.

As for EEG-based regressors, in table 7.6 the results of the LOSO validation for the BAYLEY-III scales using HRV features, are reported. For the 6-months scores the lowest MAE was reached for the cognitive one (MAE=4.8). This was obtained by a Bag regressor (Ensemble Model), with FSR as feature selection method and using the following hyperparameters:

- number of learning cycles=83; minimum leaf size=1; maximum number of splits=9 and number of features=2.

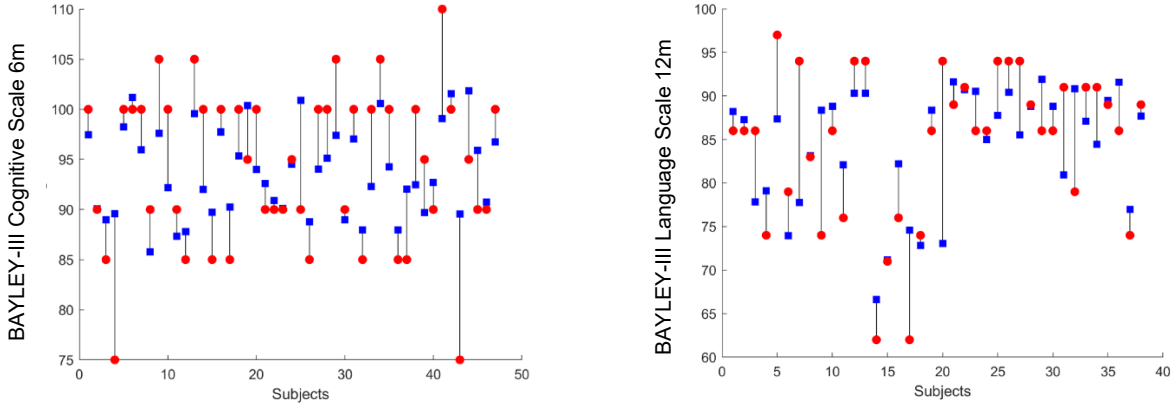
Specifically, the two features used are MSE2 (iqr) and the MDE18 (std). Note that for such scale the MAE remained lower than 5 points both for the Sepsis and the No-Sepsis group. Instead for the 12-months follow-up the best one was the language scale (Table 7.6, MAE=5.3 and R^2=0.32). This was obtained by a LsBoost regressor (Ensemble Model), with FSR as feature selection method and using the following hyperparameters:

- learn rate=0.5488, number of learning cycles=14; minimum leaf size=12; maximum number of splits=23 and number of features=2.

Specifically, the two features used are LF (mean) and the MFE12 (mean). A comparable result was achieved also for the cognitive scale (MAE=6.0 and R^2=0.27). In Figure 7.5 the predictions of the two best HRV-based models in LOSO validation are shown. In Figure 7.5a for the cognitive scores at 6-months, the predicted scores (□, blue rectangles), the true scores (○, red circles) and their respective absolute error lines (black vertical lines) are shown. The same is shown in Figure 7.4b for the Language scores at 12-months.

Table 7.6. Results of the LOSO validation for all the BAYLEY scales considered, using HRV features (m=months).

	Regressor metrics		
BAYLEY scales	**MSE**	**MAE (MAE0-MAE1)**	**R^2**
Cognitive 6m	34	4.8 (4.8 – 4.7)	0.44
Language 6m	56	5.7 (5.4 – 6.0)	0.08
Motor 6m	76	6.2 (6.6 – 5.9)	0.10
Cognitive 12m	87	6.0 (7.5 – 4.9)	0.27
Language 12m	51	5.3 (7.1 – 4.0)	0.32
Motor 12m	46	5.8 (6.1 – 5.5)	0.02



(a) (b)

Figure 7.5. (a) the predicted scores (□, blue rectangles) and the true scores (○, red circles) for the BAYLEY-III cognitive scores at 6-months using HRV features. (b) predicted and true scores for the BAYLEY-III language scores at 12-months. Vertical black lines represent the absolute error.

7.4. Discussion and Conclusions

In this work a first attempt is presented regarding the prediction by EEG- HRV-based regression models of the BAYLEY-III scores of newborns with and without sepsis at different follow-ups. The research aims at evaluating if the electroclinical characteristics of the newborn at the term age could be predictive of its neurodevelopmental scores. Moreover, this possibility was also evaluated in preterm newborns with sepsis episodes during their hospitalization, where sepsis could affect their neurodevelopment and if such alterations can be detected by AI methods. Results suggest that EEG or ECG exam, performed when the preterm newborn reached the term age, could support neonatologists or peadiatric neurologists, identifying the newborns at risk of neurodevelopmental delays.

As shown in Tables 7.5 and 7.6, promising findings suggest that, for some BAYLEY-III scales, EEG-HRV-based regression analysis could predict the neurodevelopmental scores of the newborns at 6- and 12- months. In fact, for the cognitive scale at 6-months and for language scales at 6- and 12-months, regression model were able to predict the BAYLEY-scores with a MAE lower or close to 5 points. HRV-based regression models performed better than EEG ones for all the metrics considered (MSE, MAE and R^2), also as far as the model complexity is concerned: HRV models just with two features obtained performance higher than EEG ones (which used almost 65 features). Thus HRV-based models should be preferred in terms of interpretability.

Furthermore, considering the results in Table 7.5 and 7.6, we remark that the motor scores always obtained the worst performance on all the metrics considered. This suggests that such scales might not be predictable by regression methods and thus any estimation of such scores must be considered with caution.

To the best of our knowledge, this is one of the first study that considers at the same time neurodevelopmental scores, regression analysis and sepsis. In a similar manner, Alotaibi et al. [Alotaibi et al., 2022] proposed a regression model to predict BAYLEY-III scores on newborns with hypoxic-ischemic encephalopathy. They obtained a MAE of 12.07 on cognitive scores, confirming that the regression analysis could be a valuable support tool in the neurodevelopment assessment. Furthermore, several studies confirmed that HRV features could be de predictive of neurodevelopment related scales [Thiriez et al., 2015, Leon et al., 2022, Goulding et al., 2015], although most of them considered only univariate analysis or did not consider newborns with sepsis.

The analysis of newborns with sepsis and the prediction of their BAYLEY-III scores by AI models, required the analysis of several EEG and HRV features previously used in other neonatal applications: almost 500 for EEG, and 80 for HRV, without considering the statistical descriptors expansion [Kong et al., 2018, De Wel et al., 2017, Lavanga et al., 2017, Frassineti et al., 2021a, Frassineti et al., 2021a, O'Toole et al., 2017]. This preliminary analysis led to two considerations:

- EEG-based regression models have shown lower performance both in terms of metrics and interpretability with respect to HRV-based models. This suggests that the features considered, or the pre-processing applied may be no optimal (section 7.2, e.g. the epoch length used). Thus, other EEG features have to be considered in next studies.
- In this work SVR and Ensemble models were tested. For all the scales considered Ensemble models showed better performance than SVR. Thus, Bag or LsBoost methods should be preferred for such application. However, in future studies other models should be evaluated in order to confirm it (e.g. Gaussian Process or Generalized Linear Models [Frassineti et al., 2021c]). Moreover, here FSR and ReliefF were evaluated as feature selection methods, but in future studies other methodologies could be considered such as LASSO or mRMR [Frassineti et al., 2021c].

HRV multiscale entropy features were included in the validated models. These findings confirmed that multiscale entropy indexes for HRV analysis can add helpful information in the evaluation of ANS activity in the newborn [Statello et al., 2021]. HRV entropy features were already found linked to altered cardiovascular dynamics in newborns [Frassineti et al., 2021c, Frassineti et al., 2021d, Lavanga et al., 2020]. Furthermore, for the 12-months language scales, also the Low Frequency feature (LF, statistical descriptor: mean) was included in the HRV-based regression model.

In this study, it seems that the information provided by EEG-based and HRV-based models may be seen as complementary. In fact, when just EEG or HRV was not able to predict the BAYLEY-III

score the other one might obtain reliable performance (e.g. for cognitive scale at 6-months EEG-based model obtained MAE=5.8 instead HRV-based model MAE=4.8). This suggests a possible improvement of the methods proposed in this Chapter: developing regression models able to integrate both EEG and HRV features.

We remark that there is no result in the literature about this topic, thus an optimal threshold on regression models' errors is to be defined in order to be reliably valuable in the clinical practice (ideally MSE and MAE should tend to 0 and R^2 to 1). As the sensitivity of BAYLEY-III scales is close to 5 points for our cohort, it is reasonable to consider a model as promising if it gives a MAE (or MAE1/MAE0) value lower than 5. Also notice that regression models performed better on the Sepsis group than the No-sepsis one (Tables 7.5 and 7.6, MAE1 often lower than MAE0).

Furthermore, most the newborns considered in our study had BAYLEY-III greater than the threshold usually accepted to exclude possible neurocognitive impairments (i.e. > 85-90, Table 7.3 and [Del Rosario, et al., 2020]). Therefore, our methods should be evaluated on a larger set of infants with severe impairments (i.e. with BAYLEY-III scores <75). In fact, as shown in Figures 7.4 and 7.5, the highest errors were obtained on the extreme values of the BAYLEY-III scales.

Significant differences on BAYLEY-III scores exist between sepsis and no-sepsis groups on cognitive scales at 6- and 12-months, as well as on language scales at 12-months (Table 7.3), and these differences are detectable by the regression models proposed. However, it is not possible to confirm yet if the differences are due to the preterm birth [Yiallourou et al., 2013], to sepsis or a combination of them. Anyway, even if the variations were due only to preterm birth, the models were able to predict such alterations, allowing an early detection of newborns with low BAYLEY-III scores at 6- and 12- months. Thus, further studies are required in order to confirm that sepsis affected the neurodevelopment of the infants involved, as well evaluate if sepsis might produce irreversible brain damages. In literature several studies have confirmed that sepsis may significantly alter the neurodevelopment and the proper functioning of the brain [Rallis et al., 2019]. However only few works evaluate which physiological mechanisms may occur to cause such effects. Shah et al. [Shah et al., 2008] found that sepsis may alter the white matter of newborns and that it may produce long-term effect. Thus. our results confirm that sepsis may have an active role in the neurodevelopment, but further analysis is required in order to assess the localization or the damage entity, both in the Central Nervous System and in the Autonomic Nervous System.

We remark that results are preliminary: regression models have to be validated also on 18-24 months BAYLEY-III scores to assess if and how EEG or HRV features will be able to predict scores of these scales and thus different periods of infant's neurodevelopment. At present, it can be argued that the approach proposed seems promising to predict scores of the cognitive scale at 6- months and of the language scale at 6- and 12-months. Moreover, neurodevelopmental tests, such as BAYLEY-III, being operator dependent, present several intrinsic sources of variability, limits and pitfalls and tend to overestimate the infant's development [Anderson and Burnett, 2016]. Thus, it might be challenging to obtain a perfect correspondence between models' predictions and ground-truth scores. A possible solution may be the combination of different evaluations made by several experts on the same subject. This analysis is out of the aim of the study and will be considered in future research. Moreover, the effect of possible confounding variables (e.g. BPD, ROP or NEC) could be included in future analysis to assess the results obtained about the impact of the sepsis on the neurological outcomes.

In summary our results confirmed that ANS development may be linked to CNS maturation and that alterations affecting the correct neurodevelopment of the newborn could be predicted by HRV features. If our results will be confirmed, also HRV-based regression models might be used as a

support for the clinical staff for an early detection of preterm newborns at risk of neurodevelopmental disorders. Such models could be integrated in the clinical assessment as a pre-screening test before the administration of neurodevelopmental tests planned at different follow-up periods.

Conclusions

This PhD thesis concerns the development of multimodal systems to detect and characterize neurological disorders in paediatric subjects. Specifically, the PhD work was focused on the development of Artificial Intelligence (AI) methods for: neonatal and absence seizure detection; the quantitative characterization of the speech phenotype for some genetic syndromes; the prediction of the neurodevelopmental scales in newborns with sepsis.

First attempts were made to validate EEG-based Neonatal Seizure Detectors (NSDs), showing that the neonatal seizure detection is still a tricky and time-consuming issue in the clinical practice. Heart rate variability (HRV) analysis was proposed as an alternative approach for the detection of neonatal seizures. Experimental results confirmed the involvement of the Autonomic Nervous System (ANS) during or close to neonatal seizures. Such findings were confirmed by Brain-Heart Interactions analysis. Moreover, the usefulness of multiscale entropy indexes in the characterization of neonatal seizures was shown. The comparison between EEG-based NSDs and HRV ones suggest that the best approach to detect neonatal seizures is still the EEG. However, when EEG techniques are not available, the use of HRV-based NSD could be a promising alternative. Furthermore, in future studies the possibility to merge EEG and ECG/video information into a single NSD could be evaluated. The use of EEG and/or HRV features also for the automatic classification of the aetiologies behind neonatal seizures.

Quantitative acoustical analysis has been applied for the definition of speech phenotype in genetic syndromes. Preliminary encouraging results confirm that acoustical measures could add helpful information for several syndromes with well-known language/voice impairments. Being completely non-invasive, acoustical analysis and AI methods might significantly contribute to the clinical assessment of such pathologies, also after surgical, pharmacological or logopaedic treatments and for long-term monitoring of the acoustical characteristics of the voice of these subjects. Several specific differences were found between pathological and control cases, showing a peculiar picture for each syndrome considered. Though non-optimal, the use of smartphones may allow the analysis both in a controlled environment, such as specialized clinics and hospitals, and/or in-home monitoring, after careful training of the parents or tutors of the subject.

The last part of this PhD thesis exploits the possibility of forecasting neurodevelopmental scores in preterm newborns with and without sepsis. Using AI regression models, reliable results at different time steps of the follow-up were obtained, both with EEG and HRV features. However, further studies are required in order to confirm the impact of sepsis on the neurodevelopment as related to the preterm birth, as well as if sepsis might produce irreversible brain damages. If confirmed, such methods could be used for an early detection of infants with high risk of neurodevelopment delays.

In conclusion, future studies will be devoted to the validation of such method in the clinical practice, also in prospective studies, evaluating their possible contribution to the improvement of the clinical practice and to the enhancement of the health conditions of the paediatric subjects over the years.

References

Aarabi, A., Wallois, F., Grebe, R., 2008. Does spatiotemporal synchronization of EEG change prior to absence seizures? Brain Research, Vol. 1188, pp. 207–221. https://doi.org/10.1016/j.brainres.2007.10.048.

Acharya, U.R., Vinitha Sree, S., Swapna, G., Martis, R.J., Suri, J.S., 2013. Automated EEG analysis of epilepsy: A review. Knowledge-Based Systems, Vol. 45, pp. 147–165. https://doi.org/10.1016/j.knosys.2013.02.014.

Akyüz, E., Üner, A. K., Köklü, B., Arulsamy, A., Shaikh, Mohd. F., 2021. Cardiorespiratory findings in epilepsy: A recent review on outcomes and pathophysiology. In Journal of Neuroscience Research, Vol. 99, Issue 9, pp. 2059–2073. https://doi.org/10.1002/jnr.24861.

Al-Khassaweneh, M., Villafane-Delgado, M., Mutlu, A.Y., Aviyente, S., 2016. A Measure of Multivariate Phase Synchrony Using Hyperdimensional Geometry. IEEE Trans. Signal Process, Vol. 64, Issue 11, pp. 2774–278. https://doi.org/10.1109/tsp.2016.2529586.

Alotaibi, N., Bakheet, D., Konn, D., Vollmer, B., Maharatna, K., 2022. Cognitive Outcome Prediction in Infants With Neonatal Hypoxic-Ischemic Encephalopathy Based on Functional Connectivity and Complexity of the Electroencephalography Signal. Frontiers in Human Neuroscience, Vol. 15. https://doi.org/10.3389/fnhum.2021.795006.

Alotaiby, T.N., Alshebeili, S.A., Alshawi, T., Ahmad, I., Abd El-Samie, F.E., 2014. EEG seizure detection and prediction algorithms: a survey. EURASIP J. Adv. Signal Process, Vol. 2014, Issue 1. https://doi.org/10.1186/1687-6180-2014-183.

Alshaikh, B., Yusuf, K., Sauve, R., 2013. Neurodevelopmental outcomes of very low birth weight infants with neonatal sepsis: systematic review and meta-analysis. Journal of Perinatology, Vol. 33, Issue 7, pp. 558–564. https://doi.org/10.1038/jp.2012.167.

Ames, S. G., Davis, B. S., Angus, D. C., Carcillo, J. A., Kahn, J. M., 2018. Hospital Variation in Risk-Adjusted Pediatric Sepsis Mortality. Pediatric Critical Care Medicine, Vol. 19, Issue 5, pp. 390–396. https://doi.org/10.1097/pcc.0000000000001502.

Anderson, P. J., Burnett, A., 2016. Assessing developmental delay in early childhood — concerns with the Bayley-III scales. The Clinical Neuropsychologist, Vol. 31, Issue 2, pp. 371–381. https://doi.org/10.1080/13854046.2016.1216518.

Ansari, A.H., Cherian, P.J., Caicedo, A., Naulaers, G., De Vos, M., Van Huffel, S., 2019. Neonatal Seizure Detection Using Deep Convolutional Neural Networks. Int. J. Neur. Syst., Vol. 29, Issue 04, p. 1850011. https://doi.org/10.1142/s0129065718500119.

Antonarakis, S. E., Skotko, B. G., Rafii, M. S., Strydom, A., Pape, S. E., Bianchi, D. W., Sherman, S. L., Reeves, R. H., 2020. Down syndrome. Nature reviews. Disease primers, 6(1), 9. https://doi.org/10.1038/s41572-019-0143-7.

Arias-Londoño, J. D., Godino-Llorente, J. I., Sáenz-Lechón, N., Osma-Ruiz, V., Castellanos-Domínguez, G., 2010. Automatic detection of pathological voices using complexity measures, noise parameters, and mel-cepstral coefficients. IEEE Transactions on biomedical engineering, 58(2), 370-379. doi:10.1109/TBME.2010.2089052.

Ask, T. F., Ranjitkar, S., Ulak, M., Chandyo, R. K., Hysing, M., Strand, T. A., Kvestad, I., Shrestha, L., Andreassen, M., Lugo, R. G., Shilpakar, J. S., Shrestha, M., Sütterlin, S., 2019. The Association Between Heart Rate Variability and Neurocognitive and Socio-Emotional Development in Nepalese Infants. Frontiers in Neuroscience, Vol. 13. https://doi.org/10.3389/fnins.2019.00411.

Avots, E., Jermakovs, K., Bachmann, M., Päeske, L., Ozcinar, C., Anbarjafari, G., 2022. Ensemble Approach for Detection of Depression Using EEG Features. Entropy, Vol. 24, Issue 2, p. 211. https://doi.org/10.3390/e24020211.

Awad, M., Khanna, R., 2015. Support Vector Machines for Classification. Efficient Learning Machines pp. 39–66. https://doi.org/10.1007/978-1-4302-5990-9_3.

Baboukani, S., P., Azemi, G., Boashash, B., Colditz, P., Omidvarnia, A., 2019. A novel multivariate phase synchrony measure: Application to multichannel newborn EEG analysis. Digital Signal Processing, Vol. 84, pp. 59–68. https://doi.org/10.1016/j.dsp.2018.08.019.

Balamuth, F., Weiss, S. L., Neuman, M. I., Scott, H., Brady, P. W., Paul, R., Farris, R. W. D., McClead, R., Hayes, K., Gaieski, D., Hall, M., Shah, S. S., Alpern, E. R., 2014. Pediatric Severe Sepsis in U.S. Children's Hospitals. Pediatric Critical Care Medicine, Vol. 15, Issue 9, pp. 798–805. https://doi.org/10.1097/pcc.0000000000000225.

Bandini, A., Giovannelli, F., Orlandi, S., Barbagallo, S. D., Cincotta, M., Vanni, P., Chiaramonti, R., Borgheresi, A., Zaccara, G., Manfredi, C., 2015. Automatic identification of dysprosody in idiopathic Parkinson's disease. Biomedical Signal Processing and Control, 17, 47-54. https://doi.org/10.1016/j.bspc.2014.07.006.

Bandini, A., Orlandi, S., Giovannelli, F., Felici, A., Cincotta, M., Clemente, D., Vanni, P., Zaccara, G., Manfredi, C., 2016. Markerless Analysis of Articulatory Movements in Patients With Parkinson's Disease. In Journal of Voice, Vol. 30, Issue 6, p. 766.e1-766.e11. https://doi.org/10.1016/j.jvoice.2015.10.014.

Bandt, C., Pompe, B., 2002. Permutation Entropy: A Natural Complexity Measure for Time Series. Phys. Rev. Lett, Vol. 88, Issue 17. https://doi.org/10.1103/physrevlett.88.174102.

Baumgartner, C., Koren, J.P., Rothmayer, M., 2018. Automatic Computer-Based Detection of Epileptic Seizures. Front. Neurol, Vol. 9. https://doi.org/10.3389/fneur.2018.00639.

Beck, S., Wojdyla, D., Say, L., Pilar Bertran, A., Meraldi, M., Harris Requejo, J., Rubens, C., Menon, R., Van Look, P., 2010. The worldwide incidence of preterm birth: a systematic review of maternal mortality and morbidity. In Bulletin of the World Health Organization, Vol. 88, Issue 1, pp. 31–38. https://doi.org/10.2471/blt.08.062554.

Behbahani, S., Jafarnia Dabanloo, N., Motie Nasrabadi, A., Teixeira, C. A., Dourado, A., 2014. A new algorithm for detection of epileptic seizures based on HRV signal. In Journal of Experimental & Theoretical Artificial Intelligence, Vol. 26, Issue 2, pp. 251–265. https://doi.org/10.1080/0952813x.2013.861874.

Bersani, I., Piersigilli, F., Gazzolo, D., Campi, F., Savarese, I., Dotta, A., Tamborrino, P. P., Auriti, C., Di Mambro, C., 2020. Heart rate variability as possible marker of brain damage in neonates with hypoxic ischemic encephalopathy: a systematic review. European Journal of Pediatrics, Vol. 180, Issue 5, pp. 1335–1345. https://doi.org/10.1007/s00431-020-03882-3

Blencowe, H., Cousens, S., Oestergaard, M. Z., Chou, D., Moller, A.-B., Narwal, R., Adler, A., Vera Garcia, C., Rohde, S., Say, L., Lawn, J. E., 2012. National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications. The Lancet, Vol. 379, Issue 9832, pp. 2162–2172. https://doi.org/10.1016/s0140-6736(12)60820-4.

Boersma, P., Weenink, D., 2018. Praat: doing phonetics by computer [Computer program]. Version 6.0.37, retrieved 28 August 2021 from http://www.praat.org/

Bollani, L., Ilaria Stolfi, I., Roberto Pedicino, R., Stronati, M., 2016. Manuale di Infettivologia Neonatale - II edizione. Biomedia, IIa edizione.

Bonebright, T.L., 2011. Evaluation of Auditory Display. In: Hermann, Thomas, Hunt, Andy, Neuhoff, John G. (Eds.), The Sonification Handbook. Logos Verlag, Berlin, Germany, pp. 111 –141.

Borin, A. M. S., Humeau-Heurtier, A., Murta, L. O., Silva, L. E. V., 2021. Multiscale Entropy Analysis of Short Signals: The Robustness of Fuzzy Entropy-Based Variants. Preprint (version 1) available at Research Square; 2021. https://doi.org/10.21203/rs.3.rs-361154/v1.

Boylan, G., Burgoyne, L., Moore, C., O'Flaherty, B., Rennie, J., 2010. An international survey of EEG use in the neonatal intensive care unit. Acta Paediatrica, Vol. 99, Issue 8, pp. 1150–1155. https://doi.org/10.1111/j.1651-2227.2010.01809.x.

Breakspear, M., Roberts, J.A., Terry, J.R., Rodrigues, S., Mahant, N., Robinson, P.A., 2006. A unifying explanation of primary generalized seizures through nonlinear brain modeling and bifurcation analysis. Cereb. Cortex 16, 1296–1313. https://doi.org/10.1093/cercor/bhj072.

Buchanan, G. F., 2019. Impaired CO2-Induced Arousal in SIDS and SUDEP. Trends in Neurosciences, Vol. 42, Issue 4, pp. 242–250. https://doi.org/10.1016/j.tins.2019.02.002.

Burges, C.J., 1998. A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery 2, 121–167. https://doi.org/10.1023/A:1009715923555.

Calà, F., 2022. AI techniques for the acoustic characterization of genetic syndromes. Master's Degree Thesis, date of discussion: 04/11/2022. Università degli Studi di Firenze (Firenze, Italy), Scuola di Ingegneria, Curricula Biomedical Engineering. Supervisors: Prof. Claudia Manfredi, Prof. Antonio Lanatà. Co-Supervisor: Lorenzo Frassineti.

Caliskan, A., Rencuzogullari, S., 2021. Transfer learning to detect neonatal seizure from electroencephalography signals. Neural Comput & Applic, Vol. 33, Issue 18, pp. 12087–12101. https://doi.org/10.1007/s00521-021-05878-y.

Camm, A.J., Malik, M., Bigger, J.T., Breithardt, G., Cerutti, S., Cohen, R.J., Coumel, P., Fallen, E.L., Kennedy, H.L.; Kleiger, R.E, 1996. Heart rate variability: Standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Circulation 1996, 93, 1043–1065.

Carmeli, C., Knyazeva, M.G., Innocenti, G.M., De Feo, O., 2005. Assessment of EEG synchronization based on state-space analysis. NeuroImage, Vol. 25, Issue 2, pp. 339–354. https://doi.org/10.1016/j.neuroimage.2004.11.049.

Cattani, L., Alinovi, D., Ferrari, G., Raheli, R., Pavlidis, E., Spagnoli, C., Pisani, F., 2017. Monitoring infants by automatic video processing: A unified approach to motion analysis. Computers in Biology and Medicine, Vol. 80, pp. 158–165. https://doi.org/10.1016/j.compbiomed.2016.11.010.

Chen, W., Wang, Z., Xie, H., Yu, W., 2007. Characterization of Surface EMG Signal Based on Fuzzy Entropy. IEEE Transactions on Neural Systems and Rehabilitation Engineering, Vol. 15, Issue 2, pp. 266–272. https://doi.org/10.1109/tnsre.2007.897025.

Chin, C. J., Khami, M. M., Husein, M., 2014. A general review of the otolaryngologic manifestations of Down Syndrome. International journal of pediatric otorhinolaryngology, 78(6), 899-904. https://doi.org/10.1016/j.ijporl.2014.03.012.

Chisci, L., Mavino, A., Perferi, G., Sciandrone, M., Anile, C., Colicchio, G., Fuggetta, F., 2010. Real-Time Epileptic Seizure Prediction Using AR Models and Support Vector Machines. IEEE Trans. Biomed. Eng., Vol. 57, Issue 5, pp. 1124–1132. https://doi.org/10.1109/tbme.2009.2038990.

Christodoulakis, M., Hadjipapas, A., Papathanasiou, E.S., Anastasiadou, M., Papacostas, S.S., Mitsis, G.D., 2013. On the effect of volume conduction on graph theoretic measures of brain networks in epilepsy. Modern

Electroencephalographic Assessment Techniques. Humana Press, New York, NY, pp. 103–130. https://doi.org/10.1007/7657_2013_65.

Comellas, F., Gago, S., 2007. Synchronizability of complex networks. J. Phys. A: Math. Theor., Vol. 40, Issue 17, pp. 4483–4492. https://doi.org/10.1088/1751-8113/40/17/006.

Corrales-Astorgano, M., Escudero-Mancebo, D., González-Ferreras, C., 2018. Acoustic characterization and perceptual analysis of the relative importance of prosody in speech of people with Down syndrome. Speech Communication, 99, 90-100. https://doi.org/10.1016/j.specom.2018.03.006

Costa, M., Goldberger, A. L., Peng, C.-K., 2005. Multiscale entropy analysis of biological signals. Physical Review E, Vol. 71, Issue 2. https://doi.org/10.1103/physreve.71.021906.

Costa, M., Goldberger, A., 2015. Generalized Multiscale Entropy Analysis: Application to Quantifying the Complex Volatility of Human Heartbeat Time Series. Entropy, Vol. 17, Issue 3, pp. 1197–1203. https://doi.org/10.3390/e17031197.

Costagliola, G., Orsini, A., Coll, M., Brugada, R., Parisi, P., Striano, P., 2021. The brain–heart interaction in epilepsy: implications for diagnosis, therapy, and SUDEP prevention. Annals of Clinical and Translational Neurology, Vol. 8, Issue 7, pp. 1557–1568. https://doi.org/10.1002/acn3.51382.

De Cooman, T., Varon, C., Hunyadi, B., Van Paesschen, W., Lagae, L., Van Huffel, S., 2017. Online Automated Seizure Detection in Temporal Lobe Epilepsy Patients Using Single-lead ECG. International Journal of Neural Systems, Vol. 27, Issue 07, p. 1750022. https://doi.org/10.1142/s0129065717500228.

De Wel, O., Lavanga, M., Dorado, A., Jansen, K., Dereymaeker, A., Naulaers, G., Van Huffel, S., 2017. Complexity Analysis of Neonatal EEG Using Multiscale Entropy: Applications in Brain Maturation and Sleep Stage Classification. Entropy, Vol. 19, Issue 10, p. 516. https://doi.org/10.3390/e19100516.

Deburchgraeve, W., Cherian, P.J., De Vos, M., Swarte, R.M., Blok, J.H., Visser, G.H., Govaert, P., Van Huffel, S., 2008. Automated neonatal seizure detection mimicking a human observer reading EEG. Clinical Neurophysiology, Vol. 119, Issue 11, pp. 2447–2454. https://doi.org/10.1016/j.clinph.2008.07.281.

Dejonckere, P. H., Bradley, P., Clemente, P., Cornut, G., Crevier-Buchman, L., Friedrich, G., Van De Heyning, P., Remacle, M., Woisard, V., 2001. A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. European Archives of Oto-rhino-laryngology, 258(2), 77-82. https://doi.org/10.1007/s004050000299.

Del Rosario, C., Slevin, M., Molloy, E. J., Quigley, J., Nixon, E., 2020. How to use the Bayley Scales of Infant and Toddler Development. Archives of disease in childhood - Education & practice edition, Vol. 106, Issue 2, pp. 108–112. https://doi.org/10.1136/archdischild-2020-319063.

Deller, Jr., J.H.L. Hansen, J.G. Proakis, 2000. Discrete Time Processing of Speech Signals (2d ed.), New York: IEEE Press, 2000.

Desai SS., 1997. Down syndrome: a review of the literature. Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology & Endodotics, 84:279–285. https://doi.org/10.1016/S1079-2104(97)90343-7

Dilena, R., Raviglione, F., Cantalupo, G., Cordelli, D. M., De Liso, P., Di Capua, M., Falsaperla, R., Ferrari, F., Fumagalli, M., Lori, S., Suppiej, A., Tadini, L., Dalla Bernardina, B., Mastrangelo, M., Pisani, F. (2021). Consensus protocol for EEG and amplitude-integrated EEG assessment and monitoring in neonates. In Clinical Neurophysiology Vol. 132, Issue 4, pp. 886–903. https://doi.org/10.1016/j.clinph.2021.01.012.

Ding, C., Peng, H., 2005. Minimum redundancy feature selection from microarray gene expression data. J. Bioinform. Comput. Biol., Vol. 03, Issue 02, pp. 185–205, https://doi.org/10.1142/s0219720005001004.

Dinno, A., 2015. Nonparametric pairwise multiple comparisons in independent groups using Dunn's test. The Stata Journal, 15(1), 292-300. https://doi.org/10.1177/1536867X1501500117.

Douw, L., van Dellen, E., de Groot, M., Heimans, J.J., Klein, M., Stam, C.J., Reijneveld, J.C., 2010. Epilepsy is related to theta band brain connectivity and network topology in brain tumor patients. BMC Neurosci., Vol. 11, Issue 1. https://doi.org/10.1186/1471-2202-11-103.

Doyle, O.M., Temko, A., Marnane, W., Lightbody, G., Boylan, G.B., 2010. Heart rate based automatic seizure detection in the newborn. Medical Engineering & Physics, Vol. 32, Issue 8, pp. 829–839. https://doi.org/10.1016/j.medengphy.2010.05.010.

Ekim, G., Ikizler, N., Atasoy, A., 2017. The effects of different wavelet degrees on epileptic seizure detection from EEG signals. 2017 IEEE International Conference on INnovations in Intelligent SysTems and Applications (INISTA). https://doi.org/10.1109/inista.2017.8001178.

Englert, M., Lima, L., Constantini, A. C., Latoszek, B. B. V., Maryn, Y., Behlau, M., 2019. Acoustic Voice Quality Index-AVQI para o português brasileiro: análise de diferentes materiais de fala. In CoDAS (Vol. 31). Sociedade Brasileira de Fonoaudiologia. https://doi.org/10.1590/2317-1782/20182018082.

Ershad, M., Mostafa, A., Dela Cruz, M., Vearrier, D., 2019. Neonatal Sepsis. Current Emergency and Hospital Medicine Reports, Vol. 7, Issue 3, pp. 83–90. https://doi.org/10.1007/s40138-019-00188-z

Faes, L., Marinazzo, D., Jurysta, F., Nollo, G., 2015. Linear and non-linear brain–heart and brain–brain interactions during sleep. Physiological Measurement, Vol. 36, Issue 4, pp. 683–698. https://doi.org/10.1088/0967-3334/36/4/683.

Faes, L., Marinazzo, D., Stramaglia, S., Jurysta, F., Porta, A., Giandomenico, N., 2016. Predictability decomposition detects the impairment of brain–heart dynamical networks during sleep disorders and their recovery with treatment. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, Vol. 374, Issue 2067, p. 20150177. https://doi.org/10.1098/rsta.2015.0177.

Fisher, R.S., Cross, J.H., French, J.A., Higurashi, N., Hirsch, E., Jansen, F.E., Lagae, L., Moshé, S.L., Peltola, J., Roulet Perez, E., Scheffer, I.E., Zuberi, S.M., 2017. Operational classification of seizure types by the International League Against Epilepsy: Position Paper of the ILAE Commission for Classification and Terminology. Epilepsia. https://doi.org/10.1111/epi.13670.

Fontes, A. I., Souza, P. T., Neto, A. D., Martins, A. D. M., Silveira, L. F., 2014. Classification system of pathological voices using correntropy. Mathematical Problems in Engineering, 2014. https://doi.org/10.1155/2014/924786

Fornito, A., Zalesky, A., Bullmore, E., 2016. Fundamentals of Brain Network Analysis. Academic Press.

Frassineti, L., Barba, C., Melani, F., Piras, F., Guerrini, R., Manfredi, C., 2019. Automatic detection and sonification of nonmotor generalized onset epileptic seizures: Preliminary results. Brain Research, Vol. 1721, p. 146341. https://doi.org/10.1016/j.brainres.2019.146341.

Frassineti, L., Ermini, D., Fabbri, R., Manfredi, C., 2020. Neonatal Seizures Detection using Stationary Wavelet Transform and Deep Neural Networks: Preliminary Results. 2020 IEEE 20th Mediterranean Electrotechnical Conference (MELECON). https://doi.org/10.1109/melecon48756.2020.9140713. © 2020 IEEE.

Frassineti, L., Parente, A., Manfredi, C., 2021a. Multiparametric EEG analysis of brain network dynamics during neonatal seizures. Journal of Neuroscience Methods, Vol. 348, p. 109003. https://doi.org/10.1016/j.jneumeth.2020.109003.

Frassineti, L., Lanata, A., Mandredi, C., 2021b. HRV analysis: a non-invasive approach to discriminate between newborns with and without seizures. In 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). https://doi.org/10.1109/embc46164.2021.9629741. © 2021 IEEE.

Frassineti, L., Manfredi, C., Olmi, B., Lanata, A., 2021c. A Generalized Linear Model for an ECG-based Neonatal Seizure Detector. In 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). https://doi.org/10.1109/embc46164.2021.9630841. © 2021 IEEE.

Frassineti, L., Lanatà, A., Olmi, B., Manfredi, C., 2021d. Multiscale Entropy Analysis of Heart Rate Variability in Neonatal Patients with and without Seizures. Bioengineering, Vol. 8, Issue 9, p. 122. https://doi.org/10.3390/bioengineering8090122.

Frassineti, L., Zucconi, A., Calà, F., Sforza, E., Onesimo, R., Leoni, C., Rigante, M., Manfredi, C., Zampino, G., 2021e. Analysis of vocal patterns as a diagnostic tool in patients with genetic syndromes. In 12th International Workshop, Models and Analysis of Vocal Emissions for Biomedical Applications (pp. 83-86). http://digital.casalini.it/9788855184496.

Frassineti, L., Manfredi, C., Ermini, D., Fabbri, R., Olmi, B., Lanata, A., 2022a. Analysis of Brain-Heart Interactions in newborns with and without seizures using the Convergent Cross Mapping approach. In 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). https://doi.org/10.1109/embc48229.2022.9871141. © 2022 IEEE.

Frassineti, L., Olmi, B., Lanatà, A., Lori, S., Gabbanini, S., Dani, C., Bertini, G., Cossu, C., Bastianelli, M. and Manfredi C., (2022b, accepted). Artificial Intelligence in Neonatal Intensive Care Units: a multimodal approach for seizure detection and aetiology characterization. In 66° Congresso Nazionale Società Italiana di Neurofisiologia Clinica, May 18-21, 2021, Palermo, Italia – Neurogical Sciences.

Friedman, J. H., 2001. Greedy function approximation: A gradient boosting machine. The Annals of Statistics, Vol. 29, Issue 5. https://doi.org/10.1214/aos/1013203451

Fumeaux, N. F., Ebrahim, S., Coughlin, B. F., Kadambi, A., Azmi, A., Xu, J. X., Abou Jaoude, M., Nagaraj, S. B., Thomson, K. E., Newell, T. G., Metcalf, C. S., Wilcox, K. S., Kimchi, E. Y., Moraes, M. F. D., Cash, S. S., 2020. Accurate detection of spontaneous seizures using a generalized linear model with external validation. Epilepsia, Vol. 61, Issue 9, pp. 1906–1918. https://doi.org/10.1111/epi.16628.

Gao, H.-Y., 1998. Wavelet Shrinkage Denoising Using the Non-Negative Garrote. Journal of Computational and Graphical Statistics, Vol. 7, Issue 4, pp. 469–488. https://doi.org/10.1080/10618600.1998.10474789.

Giannakopoulos, T., 2009. A method for silence removal and segmentation of speech signals, implemented in Matlab. University of Athens, Athens, 2.

González, J. J., Mañas, S., De Vera, L., Méndez, L. D., López, S., Garrido, J. M., Pereda, E. 2011. Assessment of electroencephalographic functional connectivity in term and preterm neonates. Clinical Neurophysiology, Vol. 122, Issue 4, pp. 696–702. https://doi.org/10.1016/j.clinph.2010.08.025.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2020. Generative adversarial networks. Commun. ACM. https://doi.org/10.1145/3422622.

Goodfellow, M., Rummel, C., Abela, E., Richardson, M.P., Schindler, K., Terry, J.R., 2016. Estimation of brain network ictogenicity predicts outcome from epilepsy surgery. Sci Rep., Vol. 6, Issue 1. https://doi.org/10.1038/srep29215.

Gotman, J., Flanagan, D., Zhang, J., Rosenblatt, B., 1997. Automatic seizure detection in the newborn: methods and initial evaluation. Electroencephalography and Clinical Neurophysiology, Vol. 103, Issue 3, pp. 356–362. https://doi.org/10.1016/s0013-4694(97)00003-9.

Goulding, R. M., Stevenson, N. J., Murray, D. M., Livingstone, V., Filan, P. M., Boylan, G. B., 2015. Heart rate variability in hypoxic ischemic encephalopathy: correlation with EEG grade and 2-y neurodevelopmental outcome. Pediatric Research, Vol. 77, Issue 5, pp. 681–687. https://doi.org/10.1038/pr.2015.28.

Gray, R.T., Robinson, P.A., 2009. Stability of random brain networks with excitatory and inhibitory connections. Neurocomputing 72 (7–9), 1849–1858. https://doi.org/10.1016/j.neucom.2008.06.001.

Greene, B. R., de Chazal, P., Boylan, G., Reilly, R. B., O'Brien, C., Connolly, S., 2006. Heart and respiration rate changes in the neonate during electroencephalographic seizure. Medical and Biological Engineering and Computing, Vol. 44, Issues 1–2, pp. 27–34. https://doi.org/10.1007/s11517-005-0001-5.

Greene, B.R., Boylan, G.B., Reilly, R.B., de Chazal, P., Connolly, S., 2007a. Combination of EEG and ECG for improved automatic neonatal seizure detection. Clinical Neurophysiology, Vol. 118, Issue 6, pp. 1348–1359. https://doi.org/10.1016/j.clinph.2007.02.015.

Greene, B.R., de Chazal, P., Boylan, G.B., Connolly, S., Reilly, R.B., 2007b. Electrocardiogram Based Neonatal Seizure Detection. IEEE Trans. Biomed. Eng., Vol. 54, Issue 4, pp. 673–682. https://doi.org/10.1109/tbme.2006.890137.

Gropman, A. L., Duncan, W. C., Smith, A. C., 2006. Neurologic and developmental features of the Smith-Magenis syndrome (del 17p11.2). Pediatric Neurology, 34, 337–350. https://doi.org/10.1016/j.pediatrneurol.2005.08.018.

Guimaraes, C.V.A., Donnelly, L.F., Shott, S.R., Amin, R.S., Kalra M., 2008. Relative rather than absolute macroglossia in patients with Down syndrome: implications for treatment of obstructive sleep apnea. Pediatric Radiology, 38:518–523. https://doi.org/10.1044/1092-4388(2012/12-0148).

Hämmerle, P., Eick, C., Blum, S., Schlageter, V., Bauer, A., Rizas, K. D., Eken, C., Coslovsky, M., Aeschbacher, S., Krisai, P., Meyre, P., Vesin, J., Rodondi, N., Moutzouri, E., Beer, J., Moschovitis, G., Kobza, R., Di Valentino, M., … Corino, V. D. A., 2020. Heart Rate Variability Triangular Index as a Predictor of Cardiovascular Mortality in Patients With Atrial Fibrillation. Journal of the American Heart Association, Vol. 9, Issue 15. https://doi.org/10.1161/jaha.120.016075.

Hamner, T., Udhnani, M. D., Osipowicz, K. Z., Lee, N. R., 2018. Pediatric brain development in Down syndrome: a field in its infancy. Journal of the International Neuropsychological Society, 24(9), 966-976. doi:10.1017/S1355617718000206.

Hamosh, A., et al., 2004. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Research, Vol. 33, pp. D514–D517. https://doi.org/10.1093/nar/gki033.

Hastie, T., Friedman, J., Tibshirani, R., 2001. The Elements of Statistical Learning, Springer Series in Statistics. Springer New York. https://doi.org/10.1007/978-0-387-21606-5.

He, S., Sun, K., Wang, H., 2016. Multivariate permutation entropy and its application for complexity analysis of chaotic systems. Physica A: Statistical Mechanics and its Applications, Vol. 461, pp. 812–823. https://doi.org/10.1016/j.physa.2016.06.012.

Hendrikx, D., Smits, A., Lavanga, M., De Wel, O., Thewissen, L., Jansen, K., Caicedo, A., Van Huffel, S., Naulaers, G., 2019. Measurement of Neurovascular Coupling in Neonates. Frontiers in Physiology, Vol. 10. https://doi.org/10.3389/fphys.2019.00065.

Hendrikx, D., Thewissen, L., Smits, A., Naulaers, G., Allegaert, K., Van Huffel, S., Caicedo, A., 2018. Using Graph Theory to Assess the Interaction between Cerebral Function, Brain Hemodynamics, and Systemic Variables in Premature Infants. Complexity, Vol. 2018, pp. 1–15. https://doi.org/10.1155/2018/6504039.

Hidalgo-De la Guía, I., Garayzábal-Heinze, E., Gómez-Vilda, P., Martínez-Olalla, R., Palacios-Alonso, D., 2021a. Acoustic Analysis of Phonation in Children With Smith–Magenis Syndrome. Frontiers in Human Neuroscience, 15, 661392. https://doi.org/10.3389/fnhum.2021.661392.

Hidalgo-De la Guía, I., Garayzábal, E., Gómez-Vilda, P., Palacios-Alonso, D., 2021b. Specificities of phonation biomechanics in Down syndrome children. Biomedical Signal Processing and Control, 63, 102219. https://doi.org/10.1016/j.bspc.2020.102219.

Hidalgo-De la Guía, I., Garayzábal-Heinze, E., Gómez-Vilda, P., 2020. Voice characteristics in smith–magenis syndrome: an acoustic study of laryngeal biomechanics. Languages, 5(3), 31. https://doi.org/10.3390/languages5030031.

Humeau-Heurtier, A., 2015. The Multiscale Entropy Algorithm and Its Variants: A Review. Entropy, Vol. 17, Issue 5, pp. 3110–3123. https://doi.org/10.3390/e17053110.

Humeau-Heurtier, A., 2020. Entropy Analysis in Health Informatics. Intelligent Systems Reference Library, pp. 123–143. https://doi.org/10.1007/978-3-030-54932-9_5.

Islam, M.K., Rastegarnia, A., Yang, Z., 2016a. Methods for artifact detection and removal from scalp EEG: A review. Neurophysiologie Clinique/Clinical Neurophysiology, Vol. 46, Issues 4–5, pp. 287–305. https://doi.org/10.1016/j.neucli.2016.07.002.

Islam, M.K., Rastegarnia, A., Yang, Z., 2016b. A Wavelet-Based Artifact Reduction From Scalp EEG for Epileptic Seizure Detection. IEEE J. Biomed. Health Inform, Vol. 20, Issue 5, pp. 1321–1332. https://doi.org/10.1109/jbhi.2015.2457093.

Jeppesen, J., Fuglsang-Frederiksen, A., Johansen, P., Christensen, J., Wüstenhagen, S., Tankisi, H., Qerama, E., Hess, A., Beniczky, S., 2019. Seizure detection based on heart rate variability using a wearable electrocardiography device. Epilepsia, Vol. 60, Issue 10, pp. 2105–2113. https://doi.org/10.1111/epi.16343.

Jiruska, P., de Curtis, M., Jefferys, J.G.R., Schevon, C.A., Schiff, S.J., Schindler, K., 2013. Synchronization and desynchronization in epilepsy: controversies and hypotheses. The Journal of Physiology, Vol. 591, Issue 4, pp. 787–797. https://doi.org/10.1113/jphysiol.2012.239590.

Johnson, A. E. W., Pollard, T. J., Shen, L., Lehman, L. H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Anthony Celi, L., Mark, R. G., 2016. MIMIC-III, a freely accessible critical care database. Scientific Data, Vol. 3, Issue 1. https://doi.org/10.1038/sdata.2016.35.

Karayiannis, N.B., Tao, G., Xiong, Y., Sami, A., Varughese, B., Frost, J.D., Wise, M.S., Mizrahi, E.M., 2005. Computerized Motion Analysis of Videotaped Neonatal Seizures of Epileptic Origin. Epilepsia, Vol. 46, Issue 6, pp. 901–917. https://doi.org/10.1111/j.1528-1167.2005.56504.x.

Karmakar, C., Udhayakumar, R., Palaniswami, M., 2020. Entropy Profiling: A Reduced—Parametric Measure of Kolmogorov—Sinai Entropy from Short-Term HRV Signal. Entropy, Vol. 22, Issue 12, p. 1396. https://doi.org/10.3390/e22121396.

Keilson, M.J., Allen Hauser, W., Magrill, J.P., Tepperberg, J., 1987. Ambulatory cassette EEG in absence epilepsy. Pediatric Neurology, Vol. 3, Issue 5, pp. 273–276. https://doi.org/10.1016/0887-8994(87)90067-1.

Kenta, R. D., Vorperiana, H. K., 2013. Speech Impairment in Down Syndrome: A Review. Journal of Speech, Language, and Hearing Research, 56, 178-210. https://doi.org/10.1044/1092-4388(2012/12-0148).

Kolsal, E., Serdaroğlu, A., Çilsal, E., Kula, S., Soysal, A. Ş., Kurt, A. N. Ç., Arhan, E., 2014. Can heart rate variability in children with epilepsy be used to predict seizures? Seizure, Vol. 23, Issue 5, pp. 357–362. https://doi.org/10.1016/j.seizure.2014.01.025.

Kong, A. H. T., Lai, M. M., Finnigan, S., Ware, R. S., Boyd, R. N., Colditz, P. B., 2018. Background EEG features and prediction of cognitive outcomes in very preterm infants: A systematic review. Early Human Development, Vol. 127, pp. 74–84. https://doi.org/10.1016/j.earlhumdev.2018.09.015.

König, A., Satt, A., Sorin, A., Hoory, R., Toledo-Ronen, O., Derreumaux, A., Manera, V., Verhey, F., Aalten, P., Robert, P. H., David, R., 2015. Automatic speech analysis for the assessment of patients with predementia

and Alzheimer's disease. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 1(1), 112-124. https://doi.org/10.1016/j.dadm.2014.11.012.

Kramer, G., et al., 1999. Sonification report: status of the field and research agenda. Tech. Rep., Int. Commun. Auditory Display.

Kuchenbuch, M., Benquet, P., Kaminska, A., Roubertie, A., Carme, E., de Saint Martin, A., Hirsch, E., Dubois, F., Laroche, C., Barcia, G., Chemaly, N., Milh, M., Villeneuve, N., Sauleau, P., Modolo, J., Wendling, F., Nabbout, R., 2018. Quantitative analysis and EEG markers of KCNT1 epilepsy of infancy with migrating focal seizures. Epilepsia, Vol. 60, Issue 1, pp. 20–32. https://doi.org/10.1111/epi.14605.

Lake, D. E., Richman, J. S., Griffin, M. P., Moorman, J. R., 2002. Sample entropy analysis of neonatal heart rate variability. American Journal of Physiology-Regulatory, Integrative and Comparative Physiology, Vol. 283, Issue 3, pp. R789–R797. https://doi.org/10.1152/ajpregu.00069.2002.

Lancaster, G., Iatsenko, D., Pidde, A., Ticcinelli, V., Stefanovska, A., 2018. Surrogate data for hypothesis testing of physical systems. Phys. Rep. 748, 1–60. https://doi.org/10.1016/j.physrep.2018.06.001.

Lavanga, M., De Wel, O., Caicedo, A., Jansen, K., Dereymaeker, A., Naulaers, G., Van Huffel, S., 2017. Monitoring Effective Connectivity in the Preterm Brain: A Graph Approach to Study Maturation. Complexity, Vol. 2017, pp. 1–13. https://doi.org/10.1155/2017/9078541.

Lavanga, M., De Wel, O., Caicedo, A., Jansen, K., Dereymaeker, A., Naulaers, G., Van Huffel, S., 2018. A brain-age model for preterm infants based on functional connectivity. Physiological Measurement, Vol. 39, Issue 4, p. 044006. https://doi.org/10.1088/1361-6579/aabac4.

Lavanga, M., Bollen, B., Caicedo, A., Dereymaeker, A., Jansen, K., Ortibus, E., Van Huffel, S., Naulaers, G., 2020. The effect of early procedural pain in preterm infants on the maturation of electroencephalogram and heart rate variability. Pain, Vol. 162, Issue 5, pp. 1556–1566. https://doi.org/10.1097/j.pain.0000000000002125.

Lawn, J. E., Blencowe, H., Oza, S., You, D., Lee, A. C., Waiswa, P., Lalli, M., Bhutta, Z., Barros, A. J., Christian, P., Mathers, C., Cousens, S. N., 2014. Every Newborn: progress, priorities, and potential beyond survival. The Lancet, Vol. 384, Issue 9938, pp. 189–205. https://doi.org/10.1016/s0140-6736(14)60496-7.

Lazzaro, G., Caciolo, C., Menghini, D., Cumbo, F., Digilio, M. C., Capolino, R., Zampino, G., Tartaglia, M., Vicari, S., Alfieri, P., 2020. Defining language disorders in children and adolescents with Noonan Syndrome. Molecular genetics & genomic medicine, 8(4), e1069. https://doi.org/10.1002/mgg3.1069.

Lebacq, J., Schoentgen, J., Cantarella, G., Bruss, F. T., Manfredi, C., DeJonckere, P., 2017. Maximal ambient noise levels and type of voice material required for valid use of smartphones in clinical voice research. Journal of voice, 31(5), 550-556. https://doi.org/10.1016/j.jvoice.2017.02.017.

Lee, D.-Y., Choi, Y.-S., 2018. Multiscale Distribution Entropy Analysis of Short-Term Heart Rate Variability. Entropy, Vol. 20, Issue 12, p. 952. https://doi.org/10.3390/e20120952.

Lehnertz, K., Geier, C., Rings, T., Stahn, K., 2017. Capturing time-varying brain dynamics. EPJ Nonlinear Biomed. Phys., Vol. 5, p. 2. https://doi.org/10.1051/epjnbp/2017001.

Leon, C., Cabon, S., Patural, H., Gascoin, G., Flamant, C., Roue, J.-M., Favrais, G., Beuchee, A., Pladys, P., Carrault, G., 2022. Evaluation of Maturation in Preterm Infants Through an Ensemble Machine Learning Algorithm Using Physiological Signals. IEEE Journal of Biomedical and Health Informatics, Vol. 26, Issue 1, pp. 400–410. https://doi.org/10.1109/jbhi.2021.3093096.

Li, J., Yan, J., Liu, X., Ouyang, G., 2014. Using Permutation Entropy to Measure the Changes in EEG Signals During Absence Seizures. Entropy. https://doi.org/10.3390/e16063049.

Li, P., Liu, C., Li, K., Zheng, D., Liu, C., Hou, Y., 2014. Assessing the complexity of short-term heartbeat interval series by distribution entropy. Medical & Biological Engineering & Computing, Vol. 53, Issue 1, pp. 77–87. https://doi.org/10.1007/s11517-014-1216-0.

Liu, A., Hahn, J.S., Heldt, G.P., Coen, R.W., 1992. Detection of neonatal seizures through computerized EEG analysis. Electroencephalography and Clinical Neurophysiology. https://doi.org/10.1016/0013-4694(92)90179-l.

Liu, R., Karumuri, B., Adkinson, J., Hutson, T.N., Vlachos, I., Iasemidis, L., 2018. Multivariate matching pursuit decomposition and normalized gabor entropy forquantifiation of preictal trends in epilepsy. Entropy 20 (6), 419. https://doi.org/10.3390/e20060419.

Liu, T., Yao, W., Wu, M., Shi, Z., Wang, J., Ning, X., 2017. Multiscale permutation entropy analysis of electrocardiogram. Physica A: Statistical Mechanics and its Applications, Vol. 471, pp. 492–498. https://doi.org/10.1016/j.physa.2016.11.102.

Locatelli, A., Lambicchi, L., Incerti, M., Bonati, F., Ferdico, M., Malguzzi, S., Torcasio, F., Calzi, P., Varisco, T., Paterlini, G., 2020. Is perinatal asphyxia predictable? BMC Pregnancy and Childbirth, Vol. 20, Issue 1. https://doi.org/10.1186/s12884-020-02876-1.

López-Martínez, F., Núñez-Valdez, E. R., Lorduy Gomez, J., García-Díaz, V., 2019. A neural network approach to predict early neonatal sepsis. In Computers & Electrical Engineering, Vol. 76, pp. 379–388. https://doi.org/10.1016/j.compeleceng.2019.04.015.

Loui, P., Koplin-Green, M., Frick, M., Massone, M., 2014. Rapidly Learned Identification of Epileptic Seizures from Sonified EEG. Front. Hum. Neurosci. https://doi.org/10.3389/fnhum.2014.00820.

Lucchini, M., Fifer, W. P., Sahni, R., Signorini, M. G., 2016. Novel heart rate parameters for the assessment of autonomic nervous system function in premature infants. Physiological Measurement, Vol. 37, Issue 9, pp. 1436–1446. https://doi.org/10.1088/0967-3334/37/9/1436.

Maas, A. L., Hannun, A. Y., Ng, A., Y., 2013. Rectifier nonlinearities improve neural network acoustic models. In Proc. ICML (Vol. 30, No. 1, p. 3).

Malak, J. S., Zeraati, H., Nayeri, F. S., Safdari, R., Shahraki, A. D., 2018. Neonatal intensive care decision support systems using artificial intelligence techniques: a systematic review. Artificial Intelligence Review, Vol. 52, Issue 4, pp. 2685–2704. https://doi.org/10.1007/s10462-018-9635-1.

Malarvili, M.B., Mesbah, M., 2009. Newborn Seizure Detection Based on Heart Rate Variability. IEEE Trans. Biomed. Eng. https://doi.org/10.1109/tbme.2009.2026908.

Malone, A., Anthony Ryan, C., Fitzgerald, A., Burgoyne, L., Connolly, S., Boylan, G.B., 2009. Interobserver agreement in neonatal seizure identification. Epilepsia. https://doi.org/10.1111/j.1528-1167.2009.02132.x.

Manfredi, C., Barbagallo, D., Baracca, G., Orlandi, S., Bandini, A., Dejonckere, P. H., 2015. Automatic Assessment of Acoustic Parameters of the Singing Voice: Application to Professional Western Operatic and Jazz Singers. Journal of Voice, Vol. 29, Issue 4, p. 517.e1-517.e9. https://doi.org/10.1016/j.jvoice.2014.09.014.

Manfredi, C., Lebacq, J., Cantarella, G., Schoentgen, J., Orlandi, S., Bandini, A., DeJonckere, P. H., 2017. Smartphones offer new opportunities in clinical voice research. Journal of voice, 31(1), 111-e1. https://doi.org/10.1016/j.jvoice.2015.12.020.

Manis, G., Bodini, M., Rivolta, M. W., Sassi, R., 2021. A Two-Steps-Ahead Estimator for Bubble Entropy. Entropy, Vol. 23, Issue 6, p. 761. https://doi.org/10.3390/e23060761.

Marlow, N., Wolke, D., Bracewell, M. A., Samara, M., 2005. Neurologic and Developmental Disability at Six Years of Age after Extremely Preterm Birth. New England Journal of Medicine, Vol. 352, Issue 1, pp. 9–19. https://doi.org/10.1056/nejmoa041367.

Masino, A. J., Harris, M. C., Forsyth, D., Ostapenko, S., Srinivasan, L., Bonafide, C. P., Balamuth, F., Schmatz, M., Grundmeier, R. W., 2019. Machine learning models for early sepsis recognition in the neonatal intensive care unit using readily available electronic health record data. PLOS ONE, Vol. 14, Issue 2, p. e0212665. https://doi.org/10.1371/journal.pone.0212665.

MATLAB and Statistics and Machine Learning Toolbox. The MathWorks, Inc., Natick, Massachusetts, United States.

Meijer, E. J., Hermans, K. H. M., Zwanenburg, A., Jennekens, W., Niemarkt, H. J., Cluitmans, P. J. M., van Pul, C., Wijn, P. F. F., Andriessen, P., 2014. Functional connectivity in preterm infants derived from EEG coherence analysis. European Journal of Paediatric Neurology, Vol. 18, Issue 6, pp. 780–789. https://doi.org/10.1016/j.ejpn.2014.08.003.

Mekyska, J., Janousova, E., Gomez-Vilda, P., Smekal, Z., Rektorova, I., Eliasova, I., Kostalova, M., Mrackova, M., Alonso-Hernandez, J. B., Faundez-Zanuy, M., & López-de-Ipiña, K., 2015. Robust and complex approach of pathological speech signal analysis. Neurocomputing, 167, 94-111. https://doi.org/10.1016/j.neucom.2015.02.085.

Mesbah, M., Balakrishnan, M., Colditz, P.B., Boashash, B., 2012. Automatic seizure detection based on the combination of newborn multi-channel EEG and HRV information. EURASIP J. Adv. Signal Process. https://doi.org/10.1186/1687-6180-2012-215.

Michniewicz, B., Szpecht, D., Sowińska, A., Sibiak, R., Szymankiewicz, M., Gadzinowski, J., 2020. Biomarkers in newborns with hypoxic-ischemic encephalopathy treated with therapeutic hypothermia. Child's Nervous System, Vol. 36, Issue 12, pp. 2981–2988. https://doi.org/10.1007/s00381-020-04645-z.

Midi, I., Dogan, M., Koseoglu, M., Can, G., Sehitoglu, M. A., Gunal, D. I., 2008. Voice abnormalities and their relation with motor dysfunction in Parkinson's disease. Acta Neurologica Scandinavica, 117(1), 26-34. https://doi.org/10.1111/j.1600-0404.2007.00965.x.

Mizrahi, E.M., Kellaway, P., 1987. Characterization and classification of neonatal seizures. Neurology. https://doi.org/10.1212/wnl.37.12.1837.

Molloy, E. J., Wynn, J. L., Bliss, J., Koenig, J. M., Keij, F. M., McGovern, M., Kuester, H., Turner, M. A., Giannoni, E., Mazela, J., Degtyareva, M., Strunk, T., Simons, S. H. P., Janota, J., Plotz, F. B., van den Hoogen, A., de Boode, W., Schlapbach, L. J., Reiss, I. K. M., 2020. Correction: Neonatal sepsis: need for consensus definition, collaboration and core outcomes. Pediatric Research, Vol. 90, Issue 1, pp. 232–232. https://doi.org/10.1038/s41390-020-01221-8

Monge-Álvarez, Jesús (2023). A set of Entropy measures for temporal series (1D signals) (https://www.mathworks.com/matlabcentral/fileexchange/50289-a-set-of-entropy-measures-for-temporal-series-1d-signals), MATLAB Central File Exchange. Retrieved February 24, 2023.

Moore, M. V., 1970. Speech, Hearing, and Language in de Lange Syndrome. Journal of Speech and Hearing Disorders, Vol. 35, Issue 1, pp. 66–69. https://doi.org/10.1044/jshd.3501.66.

Morelli, M. S., Orlandi, S., Manfredi, C., 2021. BioVoice: A multipurpose tool for voice analysis. Biomedical Signal Processing and Control, 64, 102302. https://doi.org/10.1016/j.bspc.2020.102302.

Morin, L., Ray, S., Wilson, C., Remy, S., Benissa, M. R., Jansen, N. J. G., Javouhey, E., Peters, M. J., Kneyber, M., De Luca, D., Nadel, S., Schlapbach, L. J., Maclaren, G., Tissieres, P., 2016. Refractory septic shock in children: a European Society of Paediatric and Neonatal Intensive Care definition. Intensive Care Medicine, Vol. 42, Issue 12, pp. 1948–1957. https://doi.org/10.1007/s00134-016-4574-2.

Mormann, F., Lehnertz, K., David, P., E. Elger, C., 2000. Mean phase coherence as a measure for phase synchronization and its application to the EEG of epilepsy patients. Physica D: Nonlinear Phenomena. https://doi.org/10.1016/s0167-2789(00)00087-7.

Mosca, F., Orfeo, L., Dani, C., 2019. Libro Bianco della Neonatologia 2019. SIN Società Italiana di Neonatologia [Online]. Available:https://www.sin-neonatologia.it/la-societa/programma-documentitriennio-2018-2021/libro-bianco-della-neonatologia-anno-2019/.

Moura, C. P., Cunha, L. M., Vilarinho, H., Cunha, M. J., Freitas, D., Palha, M., Pueschel, S. M., Pais-Clemente, M., 2008. Voice parameters in children with Down syndrome. Journal of Voice, 22(1), 34-42. https://doi.org/10.1016/j.jvoice.2006.08.011.

Mukhopadhyay, S., Puopolo, K. M., Hansen, N. I., Lorch, S. A., DeMauro, S. B., Greenberg, R. G., Cotten, C. M., Sánchez, P. J., Bell, E. F., Eichenwald, E. C., Stoll, B. J., 2020. Impact of Early-Onset Sepsis and Antibiotic Use on Death or Survival with Neurodevelopmental Impairment at 2 Years of Age among Extremely Preterm Infants. The Journal of Pediatrics, Vol. 221, pp. 39-46.e5. https://doi.org/10.1016/j.jpeds.2020.02.038.

Mutlu, A.Y., Aviyente, S., 2010. Multivariate Empirical Mode Decomposition for Quantifying Multivariate Phase Synchronization. EURASIP J. Adv. Signal Process. https://doi.org/10.1155/2011/615717.

Myers, A., Bernstein, J. A., Brennan, M.-L., Curry, C., Esplin, E. D., Fisher, J., Homeyer, M., Manning, M. A., Muller, E. A., Niemi, A.-K., Seaver, L. H., Hintz, S. R., Hudgins, L., 2014. Perinatal features of the RASopathies: Noonan syndrome, cardiofaciocutaneous syndrome and Costello syndrome. American journal of medical genetics Part A, 164(11), 2814-2821. https://doi.org/10.1002/ajmg.a.36737.

Myers, R. H., Montgomery, D. C., 1997. A Tutorial on Generalized Linear Models. Journal of Quality Technology, Vol. 29, Issue 3, pp. 274–291. https://doi.org/10.1080/00224065.1997.11979769.

Nason, G.P., Silverman, B.W., 1995. The Stationary Wavelet Transform and some Statistical Applications. Wavelets and Statistics. https://doi.org/10.1007/978-1-4612-2544-7_17.

Navakatikyan, M.A., Colditz, P.B., Burke, C.J., Inder, T.E., Richmond, J., Williams, C.E., 2006. Seizure detection algorithm for neonates based on wave-sequence analysis. Clinical Neurophysiology. https://doi.org/10.1016/j.clinph.2006.02.016.

O'Shea, A., Lightbody, G., Boylan, G., Temko, A., 2020. Neonatal seizure detection from raw multi-channel EEG using a fully convolutional architecture. Neural Networks. https://doi.org/10.1016/j.neunet.2019.11.023.

O'Shea, A., Ahmed, R., Lightbody, G., Pavlidis, E., Lloyd, R., Pisani, F., Marnane, W., Mathieson, S., Boylan, G., Temko, A., 2021. Deep Learning for EEG Seizure Detection in Preterm Infants. Int. J. Neur. Syst. https://doi.org/10.1142/s0129065721500088.

O'Toole, J.M., Boylan, G.B., Lloyd, R.O., Goulding, R.M., Vanhatalo, S., Stevenson, N.J., 2017. Detecting bursts in the EEG of very and extremely premature infants using a multi-feature approach. Med. Eng. Phys. 45, 42–50. https://doi.org/10.1016/j.medengphy.2017.04.003.

Olmi, B., Frassineti, L., Lanata, A., Manfredi, C., 2021. Automatic Detection of Epileptic Seizures in Neonatal Intensive Care Units Through EEG, ECG and Video Recordings: A Survey. IEEE Access. https://doi.org/10.1109/access.2021.3118227. © 2021 IEEE.

Olmi, B., Manfredi, C., Frassineti, L., Dani, C., Lori, S., Bertini, G., Cossu, C., Bastianelli, M., Gabbanini, S., Lanatà, A., 2022a. Heart Rate Variability Analysis for Seizure Detection in Neonatal Intensive Care Units. Bioengineering. https://doi.org/10.3390/bioengineering9040165.

Olmi, B., Manfredi, C., Frassineti, L., Dani, C., Lori, S., Bertini, G., Gabbanini, S., Lanata, A., 2022b. Aggregate Channel Features for newborn face detection in Neonatal Intensive Care Units. 2022 44th Annual

International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). https://doi.org/10.1109/embc48229.2022.9871399. © 2022 IEEE.

Omidvarnia, A., Azemi, G., Colditz, P.B., Boashash, B., 2013. A time–frequency based approach for generalized phase synchrony assessment in nonstationary multivariate signals. Digital Signal Processing. https://doi.org/10.1016/j.dsp.2013.01.002.

Ortgies, T., Rullmann, M., Ziegelhöfer, D., Bläser, A., Thome, U. H., 2021. The role of early-onset-sepsis in the neurodevelopment of very low birth weight infants. BMC Pediatrics, Vol. 21, Issue 1. https://doi.org/10.1186/s12887-021-02738-5.

Oshima, K., Carmeli, C., Hasler, M., 2006. State change detection using multivariate synchronization measure from physiological signals. Journal of Signal Processing, 10, 223-226.

Pan, J., Tompkins, W. J., 1985. A Real-Time QRS Detection Algorithm. IEEE Transactions on Biomedical Engineering, Vol. BME-32 Issue 3, pp. 230–236. https://doi.org/10.1109/tbme.1985.325532.

Parente, A., 2022. Machine learning techniques for studying the influence of sepsis on neurodevelopment in preterm infants: application to the Neonatal Intensive Care Unit AOU Careggi, Florence. Master's Degree Thesis, date of discussion: 07/21/2022. Università degli Studi di Firenze (Firenze, Italy), Scuola di Ingegneria, Curricula Biomedical Engineering. Supervisors: Prof. Claudia Manfredi, Prof. Antonio Lanatà. Co-Supervisor: Lorenzo Frassineti.

Patel, A.X., Kundu, P., Rubinov, M., Jones, P.S., Vértes, P.E., Ersche, K.D., Suckling, J., Bullmore, E.T., 2014. A wavelet method for modeling and despiking motion artifacts from resting-state fMRI time series. NeuroImage. https://doi.org/10.1016/j.neuroimage.2014.03.012.

Paul, Y., 2018. Various epileptic seizure detection techniques using biomedical signals: a review. Brain Inf. https://doi.org/10.1186/s40708-018-0084-z.

Pavel, A.M., Rennie, J.M., de Vries, L.S., Blennow, M., Foran, A., Shah, D.K., Pressler, R.M., Kapellou, O., Dempsey, E.M., Mathieson, S.R., Pavlidis, E., van Huffelen, A.C., Livingstone, V., Toet, M.C., Weeke, L.C., Finder, M., Mitra, S., Murray, D.M., Marnane, W.P., Boylan, G.B., 2020. A machine-learning algorithm for neonatal seizure recognition: a multicentre, randomised, controlled trial. The Lancet Child & Adolescent Health. https://doi.org/10.1016/s2352-4642(20)30239-x.

Pek, J. H., Yap, B. J., Gan, M. Y., Seethor, S. T. T., Greenberg, R., Hornik, C. P. V., Tan, B., Lee, J. H., Chong, S.-L., 2020. Neurocognitive impairment after neonatal sepsis: protocol for a systematic review and meta-analysis. BMJ Open, Vol. 10, Issue 6, p. e038816. https://doi.org/10.1136/bmjopen-2020-038816.

Penney, J., Gibson, A., Cox, F., Proctor, M., Szakay, A., 2021. A comparison of acoustic correlates of voice quality across different recording devices: A cautionary tale. In 22nd Annual Conference of the International Speech Communication Association, INTERSPEECH 2021 (pp. 4845-4849). doi: 10.21437/Interspeech.2021-729

Perrella, A., Sorelli, M., Giardini, F., Frassineti, L., Francia, P., Bocchi, L., 2018. Wavelet Phase Coherence Between the Microvascular Pulse Contour and the Respiratory Activity. In IFMBE Proceedings, pp. 311–314. https://doi.org/10.1007/978-981-10-9038-7_58.

Pierpont, E. I., Weismer, S. E., Roberts, A. E., Tworog-Dube, E., Pierpont, M. E., Mendelsohn, N. J., Seidenberg, M. S., 2010. The language phenotype of children and adolescents with Noonan syndrome. Journal of speech, language, and hearing research: JSLHR, 53(4), 917-932. doi: 10.1044/1092-4388(2009/09-0046).

Pincus, S. M., 1991. Approximate entropy as a measure of system complexity. In Proceedings of the National Academy of Sciences, Vol. 88, Issue 6, pp. 2297–2301 https://doi.org/10.1073/pnas.88.6.2297.

Pisani, F., Spagnoli, C., Pavlidis, E., Facini, C., Kouamou Ntonfo, G.M., Ferrari, G., Raheli, R., 2014. Real-time automated detection of clonic seizures in newborns. Clinical Neurophysiology. https://doi.org/10.1016/j.clinph.2013.12.119.

Ponten, S.C., Bartolomei, F., Stam, C.J., 2007. Small-world networks and epilepsy: graph theoretical analysis of intracerebrally recorded mesial temporal lobe seizures. Clin. Neurophysiol. 118 (4), 918–927. https://doi.org/10.1016/j.clinph.2006.12.002.

Porta, A., Bari, V., Marchi, A., De Maria, B., Castiglioni, P., di Rienzo, M., Guzzetti, S., Cividjian, A., Quintin, L., 2015. Limits of permutation-based entropies in assessing complexity of short heart period variability. Physiological Measurement, Vol. 36, Issue 4, pp. 755–765. https://doi.org/10.1088/0967-3334/36/4/755.

Pressler, R.M., Lagae, L., 2020. Why we urgently need improved seizure and epilepsy therapies for children and neonates. Neuropharmacology. https://doi.org/10.1016/j.neuropharm.2019.107854.

Pressler, R.M., Cilio, M.R., Mizrahi, E.M., Moshé, S.L., Nunes, M.L., Plouin, P., Vanhatalo, S., Yozawitz, E., de Vries, L.S., Puthenveettil Vinayan, K., Triki, C.C., Wilmshurst, J.M., Yamamoto, H., Zuberi, S.M., 2021. The ILAE classification of seizures and the epilepsies: Modification for seizures in the neonate. Position paper by the ILAE Task Force on Neonatal Seizures. Epilepsia. https://doi.org/10.1111/epi.16815.

Prout, A. J., Talisa, V. B., Carcillo, J. A., Mayr, F. B., Angus, D. C., Seymour, C. W., Chang, C.-C. H., Yende, S., 2018. Children with Chronic Disease Bear the Highest Burden of Pediatric Sepsis. The Journal of Pediatrics, Vol. 199, pp. 194-199.e1. https://doi.org/10.1016/j.jpeds.2018.03.056.

Qian, N., 1999. On the momentum term in gradient descent learning algorithms. Neural Networks. https://doi.org/10.1016/s0893-6080(98)00116-6.

Quiroga, R.Q., Kraskov, A., Kreuz, T., Grassberger, P., 2002. Performance of different synchronization measures in real data: a case study on electroencephalographic signals. Phys. Rev. E 65 (4), 041903. https://doi.org/10.1103/PhysRevE.65.041903.

Rallis, D., Karagianni, P., Goutsiou, E., Soubasi-Griva, V., Banerjee, J., Tsakalidis, C., 2019. The association of the cerebral oxygenation during neonatal sepsis with the Bayley-III Scale of Infant and Toddler Development index scores at 18–24 months of age. Early Human Development, Vol. 136, pp. 49–53. https://doi.org/10.1016/j.earlhumdev.2019.07.008.

Ramantani, G., Schmitt, B., Plecko, B., Pressler, R.M., Wohlrab, G., Klebermass-Schrehof, K., Hagmann, C., Pisani, F., Boylan, G.B., 2019. Neonatal Seizures—Are We there Yet? Neuropediatrics. https://doi.org/10.1055/s-0039-1693149.

Rankine, L., Stevenson, N., Mesbah, M., Boashash, B., 2006. A nonstationary model of newborn EEG. IEEE Trans. Biomed. Eng. 54 (1), 19–28. https://doi.org/10.1109/TBME.2006.886667.

Räsänen, O., Metsäranta, M., Vanhatalo, S., 2013. Development of a novel robust measure for interhemispheric synchrony in the neonatal EEG: Activation Synchrony Index (ASI). NeuroImage, Vol. 69, pp. 256–266. https://doi.org/10.1016/j.neuroimage.2012.12.017.

Redelico, F., Traversaro, F., García, M., Silva, W., Rosso, O., Risk, M., 2017. Classification of Normal and Pre-Ictal EEG Signals Using Permutation Entropies and a Generalized Linear Model as a Classifier. Entropy, Vol. 19, Issue 2, p. 72. https://doi.org/10.3390/e19020072.

Ricci Maccarini, A., Lucchini, E., 2002. La valutazione soggettiva ed oggettiva della disfonia. Il protocollo SIFEL in “Relazione Ufficiale al XXIV Congresso Nazionale della Società Italiana di Foniatria e Logopedia. Acta Phon. Lat, 26, 1-2.

Richman, J. S., Moorman, J. R., 2000. Physiological time-series analysis using approximate entropy and sample entropy. American Journal of Physiology-Heart and Circulatory Physiology, Vol. 278, Issue 6, pp. H2039–H2049. https://doi.org/10.1152/ajpheart.2000.278.6.h2039.

Roberts AE. Noonan Syndrome. 2001 Nov 15 [Updated 2022 Feb 17]. In: Adam MP, Everman DB, Mirzaa GM, et al., editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2022. Available from: https://www.ncbi.nlm.nih.gov/books/NBK1124/.

Robnik-Šikonja, M., Kononenko, I., 2003. Theoretical and Empirical Analysis of ReliefF and RReliefF. Machine Learning 53, 23–69. https://doi.org/10.1023/A:1025667309714.

Ropper, A., H., Adams, D., R., Victor, M., Brown, H., R., 2005. Adams and Victor's principles of neurology. McGraw-Hill Medical Pub. Division, New York.

Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T.H., Faubert, J., 2019. Deep learning-based electroencephalography analysis: a systematic review. J. Neural Eng. https://doi.org/10.1088/1741-2552/ab260c.

Sachs, J.E., Berkovits, S. 1984. Probabilistic Analysis and Performance Modelling of the 'Swedish' Algorithm and Modifications. In: Chaum, D. (eds) Advances in Cryptology. Springer, Boston, MA. https://doi.org/10.1007/978-1-4684-4730-9_21.

Scher, M. S., 2021. "The First Thousand Days" Define a Fetal/Neonatal Neurology Program. Frontiers in Pediatrics, Vol. 9. https://doi.org/10.3389/fped.2021.683138.

Schiecke, K., Pester, B., Piper, D., Benninger, F., Feucht, M., Leistritz, L., Witte, H., 2016. Nonlinear Directed Interactions Between HRV and EEG Activity in Children With TLE. IEEE Transactions on Biomedical Engineering, Vol. 63, Issue 12, pp. 2497–2504. https://doi.org/10.1109/tbme.2016.2579021.

Schiecke, K., Schumann, A., Benninger, F., Feucht, M., Baer, K.-J., Schlattmann, P., 2019. Brain–heart interactions considering complex physiological data: processing schemes for time-variant, frequency-dependent, topographical and statistical examination of directed interactions by convergent cross mapping. Physiological Measurement, Vol. 40, Issue 11, p. 114001. https://doi.org/10.1088/1361-6579/ab5050.

Schindler, K., Leung, H., Elger, C.E., Lehnertz, K., 2006. Assessing seizure dynamics by analysing the correlation structure of multichannel intracranial EEG. Brain. https://doi.org/10.1093/brain/awl304.

Schindler, K.A., Bialonski, S., Horstmann, M.-T., Elger, C.E., Lehnertz, K., 2008. Evolving functional network properties and synchronizability during human epileptic seizures. Chaos. https://doi.org/10.1063/1.2966112.

Serrano, M.A., Boguna, M., Vespignani, A., 2009. Extracting the multiscale backbone of complex weighted networks. Proc. Natl. Acad. Sci. U.S.A. 106 (16), 6483–6488. https://doi.org/10.1073/pnas.0808904106.

Shaffer, F., Ginsberg, J. P., 2017. An Overview of Heart Rate Variability Metrics and Norms. Frontiers in Public Health, Vol. 5. https://doi.org/10.3389/fpubh.2017.00258.

Shah, D. K., Doyle, L. W., Anderson, P. J., Bear, M., Daley, A. J., Hunt, R. W., Inder, T. E., 2008. Adverse Neurodevelopment in Preterm Infants with Postnatal Sepsis or Necrotizing Enterocolitis is Mediated by White Matter Abnormalities on Magnetic Resonance Imaging at Term. The Journal of Pediatrics, Vol. 153, Issue 2, pp. 170-175.e1. https://doi.org/10.1016/j.jpeds.2008.02.033.

Shellhaas, R.A., 2019. Seizure classification, etiology, and management. Handbook of Clinical Neurology. https://doi.org/10.1016/b978-0-444-64029-1.00017-5.

Shellhaas, R.A., Soaita, A.I., Clancy, R.R., 2007. Sensitivity of Amplitude-Integrated Electroencephalography for Neonatal Seizure Detection. Pediatrics. https://doi.org/10.1542/peds.2007-0514.

Shorten, C., Khoshgoftaar, T.M., 2019. A survey on Image Data Augmentation for Deep Learning. J Big Data. https://doi.org/10.1186/s40537-019-0197-0.

Silvani, A., Calandra-Buonaura, G., Dampney, R. A. L., Cortelli, P., 2016. Brain–heart interactions: physiology and clinical implications. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, Vol. 374, Issue 2067, p. 20150181. https://doi.org/10.1098/rsta.2015.0181.

Singer, M., Deutschman, C. S., Seymour, C. W., Shankar-Hari, M., Annane, D., Bauer, M., Bellomo, R., Bernard, G. R., Chiche, J.-D., Coopersmith, C. M., Hotchkiss, R. S., Levy, M. M., Marshall, J. C., Martin, G. S., Opal, S. M., Rubenfeld, G. D., van der Poll, T., Vincent, J.-L., Angus, D. C., 2016. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA, Vol. 315, Issue 8, p. 801. https://doi.org/10.1001/jama.2016.0287

Smith, K., Spyrou, L., Escudero, J., 2019. Graph-Variate Signal Analysis. IEEE Trans. Signal Process. https://doi.org/10.1109/tsp.2018.2881658.

Song, W., Jung, S. Y., Baek, H., Choi, C. W., Jung, Y. H., Yoo, S., 2020. A Predictive Model Based on Machine Learning for the Early Detection of Late-Onset Neonatal Sepsis: Development and Observational Study. JMIR Medical Informatics, Vol. 8, Issue 7, p. e15965. https://doi.org/10.2196/15965.

Stam, C.J., Nolte, G., Daffertshofer, A., 2007. Phase lag index: Assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources. Hum. Brain Mapp. https://doi.org/10.1002/hbm.20346.

Statello, R., Carnevali, L., Alinovi, D., Pisani, F., Sgoifo, A., 2018. Heart rate variability in neonatal patients with seizures. Clinical Neurophysiology, Vol. 129, Issue 12, pp. 2534–2540. https://doi.org/10.1016/j.clinph.2018.10.001.

Statello, R., Carnevali, L., Sgoifo, A., Miragoli, M., Pisani, F., 2021. Heart rate variability in neonatal seizures: Investigation and implications for management. Neurophysiologie Clinique, Vol. 51, Issue 6, pp. 483–492. https://doi.org/10.1016/j.neucli.2021.10.002.

Stevenson, N.J., Tapani, K., Lauronen, L., Vanhatalo, S., 2019. A dataset of neonatal EEG recordings with seizure annotations. Sci Data. https://doi.org/10.1038/sdata.2019.39.

Stojanovik, V., 2021. Genetic syndromes and communication disorders. The handbook of language and speech disorders, 95-109. https://doi.org/10.1002/9781119606987.ch5.

Sugihara, G., May, R., Ye, H., Hsieh, C., Deyle, E., Fogarty, M., Munch, S., 2012. Detecting Causality in Complex Ecosystems. Science, Vol. 338, Issue 6106, pp. 496–500. https://doi.org/10.1126/science.1227079.

Sun, G., Fan, Z., Mastorakis, N. E., Kaminaris, S. D., Zhuang, X., 2017. The complexity analysis of voiced and unvoiced speech signal based on sample entropy. In 2017 Fourth International Conference on Mathematics and Computers in Sciences and in Industry (MCSI) (pp. 26-29). IEEE. doi: 10.1109/MCSI.2017.14.

Tanveer, M.A., Khan, M.J., Sajid, H., Naseer, N., 2021. Convolutional neural networks ensemble model for neonatal seizure detection. Journal of Neuroscience Methods. https://doi.org/10.1016/j.jneumeth.2021.109197.

Tapani, K.T., Vanhatalo, S., Stevenson, N.J., 2019. Time-Varying EEG Correlations Improve Automated Neonatal Seizure Detection. Int. J. Neur. Syst. https://doi.org/10.1142/s0129065718500302.

Tarvainen, M. P., Niskanen, J.-P., Lipponen, J. A., Ranta-aho, P. O., Karjalainen, P. A., 2014. Kubios HRV – Heart rate variability analysis software. Computer Methods and Programs in Biomedicine, Vol. 113, Issue 1, pp. 210–220. https://doi.org/10.1016/j.cmpb.2013.07.024.

Teixeira, J. P., Oliveira, C., Lopes, C., 2013. Vocal acoustic analysis–jitter, shimmer and hnr parameters. Procedia Technology, 9, 1112-1122. https://doi.org/10.1016/j.protcy.2013.12.124.

Temko, A., Thomas, E., Marnane, W., Lightbody, G., Boylan, G.B., 2011a. Performance assessment for EEG-based neonatal seizure detectors. Clinical Neurophysiology. https://doi.org/10.1016/j.clinph.2010.06.035.

Temko, A., Thomas, E., Marnane, W., Lightbody, G., Boylan, G., 2011b. EEG-based neonatal seizure detection with Support Vector Machines. Clinical Neurophysiology. https://doi.org/10.1016/j.clinph.2010.06.034.

Temko, A., Marnane, W., Boylan, G., Lightbody, G., 2015a. Clinical implementation of a neonatal seizure detection algorithm. Decision Support Systems. https://doi.org/10.1016/j.dss.2014.12.006

Temko, A., Doyle, O., Murray, D., Lightbody, G., Boylan, G., Marnane, W., 2015b. Multimodal predictor of neurodevelopmental outcome in newborns with hypoxic-ischaemic encephalopathy. Computers in Biology and Medicine. https://doi.org/10.1016/j.compbiomed.2015.05.017.

Temko, A., Lightbody, G., 2016. Detecting Neonatal Seizures With Computer Algorithms. Journal of Clinical Neurophysiology. https://doi.org/10.1097/wnp.0000000000000295.

Temko, A., Sarkar, A.Kr., Boylan, G.B., Mathieson, S., Marnane, W.P., Lightbody, G., 2017. Toward a Personalized Real-Time Diagnosis in Neonatal Seizure Detection. IEEE J. Transl. Eng. Health Med. https://doi.org/10.1109/jtehm.2017.2737992.

Thakur, J., Pahuja, S., Pahuja, R., 2017. Performance Comparison of Systemic Inflammatory Response Syndrome with Logistic Regression Models to Predict Sepsis in Neonates. Children, Vol. 4, Issue 12, p. 111. https://doi.org/10.3390/children4120111.

Thakur, J., Pahuja, S. K., Pahuja, R., 2020. Temperature as a Predictor of Neonatal Sepsis. Advances in Intelligent Systems and Computing pp. 1373–1379. https://doi.org/10.1007/978-981-15-0751-9_125.

Thibeault-Eybalin, M.-P., Lortie, A., Carmant, L., 2009. Neonatal Seizures: Do They Damage the Brain? Pediatric Neurology. https://doi.org/10.1016/j.pediatrneurol.2008.10.026.

Thiriez, G., Mougey, C., Vermeylen, D., Wermenbol, V., Lanquart, J.-P., Lin, J. S., Franco, P., 2015. Altered autonomic control in preterm newborns with impaired neurological outcomes. Clinical Autonomic Research, Vol. 25, Issue 4, pp. 233–242. https://doi.org/10.1007/s10286-015-0298-6.

Thomas, E.M., Temko, A., Lightbody, G., Marnane, W.P., Boylan, G.B., 2010. Gaussian mixture models for classification of neonatal seizures using EEG. Physiol. Meas. https://doi.org/10.1088/0967-3334/31/7/013.

Tokariev, A., Stjerna, S., Lano, A., Metsäranta, M., Palva, J.M., Vanhatalo, S., 2018. Preterm Birth Changes Networks of Newborn Cortical Activity. Cerebral Cortex. https://doi.org/10.1093/cercor/bhy100.

Tong, S., Koller, D., 2002. Support vector machine active learning with applications to text classification. Journal of Machine Learning Research, 2(1), 45–66. https://doi.org/10.1162/153244302760185243.

Toole, J. M. O., Boylan, G. B. (2017). NEURAL: quantitative features for newborn EEG using Matlab (Version 1). arXiv. https://doi.org/10.48550/ARXIV.1704.05694.

Tóth, B., Urbán, G., Háden, G.P., Márk, M., Török, M., Stam, C.J., Winkler, I., 2017. Large-scale network organization of EEG functional connectivity in newborn infants. Hum. Brain Mapp. https://doi.org/10.1002/hbm.23645.

Tressoldi, P. E., Stella, G., Faggella, M., 2001. The development of reading speed in Italians with dyslexia: A longitudinal study. Journal of learning disabilities, 34(5), 414-417. https://doi.org/10.1177/002221940103400503.

Tsuchida, T.N., Wusthoff, C.J., Shellhaas, R.A., Abend, N.S., Hahn, C.D., Sullivan, J.E., Nguyen, S., Weinstein, S., Scher, M.S., Riviello, J.J., Clancy, R.R., 2013. American Clinical Neurophysiology Society

Standardized EEG Terminology and Categorization for the Description of Continuous EEG Monitoring in Neonates. Journal of Clinical Neurophysiology. https://doi.org/10.1097/wnp.0b013e3182872b24.

Unicef Report 2019. Levels & Trends in Child Mortality. Report 2019: estimates developed by the UN Interagency Group for Child Mortality Estimation.

Upadhyay, R., Padhy, P.K., Kankar, P.K., 2016. A comparative study of feature ranking techniques for epileptic seizure detection using wavelet transform. Computers & Electrical Engineering. https://doi.org/10.1016/j.compeleceng.2016.05.016.

Valenza, G., Garcia, R. G., Citi, L., Scilingo, E. P., Tomaz, C. A., Barbieri, R., 2015. Nonlinear digital signal processing in mental health: characterization of major depression using instantaneous entropy measures of heartbeat dynamics. Frontiers in Physiology, Vol. 6. https://doi.org/10.3389/fphys.2015.00074.

Valenza, G., Greco, A., Gentili, C., Lanata, A., Sebastiani, L., Menicucci, D., Gemignani, A., Scilingo, E. P., 2016. Combining electroencephalographic activity and instantaneous heart rate for assessing brain–heart dynamics during visual emotional elicitation in healthy subjects. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, Vol. 374, Issue 2067, p. 20150176. https://doi.org/10.1098/rsta.2015.0176.

van Diessen, E., Diederen, S.J.H., Braun, K.P.J., Jansen, F.E., Stam, C.J., 2013. Functional and structural brain networks in epilepsy: What have we learned? Epilepsia. https://doi.org/10.1111/epi.12350.

van Putten, M. J. A. M., 2007. The revised brain symmetry index. Clinical Neurophysiology, Vol. 118, Issue 11, pp. 2362–2367. https://doi.org/10.1016/j.clinph.2007.07.019.

van Rooij, L.G.M., Hellström-Westas, L., de Vries, L.S., 2013. Treatment of neonatal seizures. Seminars in Fetal and Neonatal Medicine. https://doi.org/10.1016/j.siny.2013.01.001.

Verrotti, A., Matricardi, S., Rinaldi, V.E., Prezioso, G., Coppola, G., 2015. Neuropsychological impairment in childhood absence epilepsy: Review of the literature. Journal of the Neurological Sciences. https://doi.org/10.1016/j.jns.2015.10.035.

Vidyaratne, L.S., Iftekharuddin, K.M., 2017. Real-Time Epileptic Seizure Detection Using EEG. IEEE Trans. Neural Syst. Rehabil. Eng. https://doi.org/10.1109/tnsre.2017.2697920.

Volpe, J.J., 1989. Neonatal Seizures: Current Concepts and Revised Classification. Pediatrics. https://doi.org/10.1542/peds.84.3.422.

Wang, Z., Zhang, P., Zhou, W., Zhou, X., Shi, Y., Cheng, X., Lin, Z., Xia, S., Zhou, W., & Cheng, G., 2021. Electroencephalography monitoring in the neonatal intensive care unit: a Chinese perspective. Translational Pediatrics, Vol. 10, Issue 3, pp. 552–559. https://doi.org/10.21037/tp-20-340.

Weiss, S. L., Balamuth, F., Hensley, J., Fitzgerald, J. C., Bush, J., Nadkarni, V. M., Thomas, N. J., Hall, M., Muszynski, J., 2017. The Epidemiology of Hospital Death Following Pediatric Severe Sepsis. In Pediatric Critical Care Medicine, Vol. 18, Issue 9, pp. 823–830. https://doi.org/10.1097/pcc.0000000000001222.

Weiss, S. L., Peters, M. J., et al., 2020. Surviving sepsis campaign international guidelines for the management of septic shock and sepsis-associated organ dysfunction in children. Intensive Care Medicine, Vol. 46, Issue S1, pp. 10–67. https://doi.org/10.1007/s00134-019-05878-6.

World Health Organization 2012; March of Dimes; The Partnership for Maternal, Newborn & Child Health; Save the Children. Born too soon: the global action report on preterm birth. Born too soon.

Yiallourou, S. R., Witcombe, N. B., Sands, S. A., Walker, A. M., Horne, R. S. C., 2013. The development of autonomic cardiovascular control is altered by preterm birth. Early Human Development, Vol. 89, Issue 3, pp. 145–152. https://doi.org/10.1016/j.earlhumdev.2012.09.009.

Yıldız, G. U., Dogan, E. A., Dogan, U., Tokgoz, O. S., Ozdemir, K., Genc, B. O., İlhan, N., 2011. Analysis of 24-hour heart rate variations in patients with epilepsy receiving antiepileptic drugs. Epilepsy & Behavior, Vol. 20, Issue 2, pp. 349–354. https://doi.org/10.1016/j.yebeh.2010.12.001.

Zeng, K., Yan, J., Wang, Y., Sik, A., Ouyang, G., Li, X., 2016. Automatic detection of absence seizures with compressive sensing EEG. Neurocomputing. https://doi.org/10.1016/j.neucom.2015.06.07

Acknowledgements

It was a long, amazing and unpredictable journey.

First of all, I am very grateful to my tutor Prof. Claudia Manfredi, for all the opportunities, motivations, inspirations, suggestions and encouragements that she gave me during all these years. She taught me that with perseverance and patience everything is possible. Moreover, thanks to Prof. Antonio Lanatà for his relevant support in this challenging era of COVID-19, hoping that this collaboration could continue in the next years.

Thanks to all the GenOMeC Faculty Board, especially to the coordinator Prof. Francesca Ariani, to Prof. Alessandra Renieri and to my supervisor Prof. Simone Furini. Furthermore, thanks to Dr. Rita Paradiso and all the SmarTex team, for allowing me to collaborate with them in their stimulating and interesting activities.

A special thanks to all the Careggi team, in particular to Dr. Silvia Lori and Dr. Giovanna Bertini who believed in my research activity, dedicating time and efforts to support me. I would also like to thank Dr. Simonetta Gabbanini, Dr. Cesarina Cossu, Dr. Maria Bastianelli and Dr. Clara Lunardi for all the time they dedicated to me, and for collecting the Careggi Dataset and the BAYLEY-III Dataset. Moreover thanks to Prof. Carlo Dani for the coordination of the research activities between our laboratory and the Careggi team.

Thanks to the A. Meyer team, where my interest in neonatal and paediatric research began with my Master's Degree Thesis. In particular, I would like to thank Dr. Carmen Barba, Dr. Federico Melani, Dr. Francesca Piras, Eng. Matteo Lenge and Prof. Renzo Guerrini. I will never forget the time that I spent with you.

Thanks to all the ex-students and researchers of the Biomedical Engineering Laboratory: Daniele, Martino, Valentina V., Sara (thank you for your help in Glasgow), Eva, Cosimo, Eleonora B., Federico and Pietro. Especially, thanks to Eng. Benedetta Olmi, Eng. Rachele Fabbri and Eng. Angela Parente for the patience and for the trust they have placed in me and I in them. I'm sure that in the future our research paths are bound to cross again.

Thanks to my friends: Lorenzo G., Matteo, Emanuele, Simone and Cecilia. Thanks to Francesco G., Benedetta, Celeste and last but not least Valentina G. who built a bridge between my two worlds. Thanks to my ex-colleagues and friends of the University of Florence: Duccio, Dino, Eleonora T., Lorenzo F., Luca, Filippo, Jack, Marco, Mattia and Francesco. We are still in contact, and our periodic reunions are the proof that our bond is still vivid and solid after years from our graduation.

Thanks to my family for her support during all these years. Thanks to my parents and my grandfather Livio with all my heart, who inspired me and instilled in me the passion for science and research.

It was a long, amazing and unpredictable journey, and I am grateful to have shared it with you.

And finally, Silvia, the last words are for you, but no words could completely explain what it means to have you next to me. Without you, nothing could have been as it is now.