NMR Spectroscopy Combined with Machine Learning Approaches for Age Prediction in Healthy and Parkinson's Disease Cohorts through Metabolomic Fingerprints

Giovanna Maria Dimitri [1,2,†], Gaia Meoni [3,4,†], Leonardo Tenori [3,4], Claudio Luchinat [3,4,*] and Pietro Lió [2,*,‡] on behalf of the PROPAG-AGEING Consortium

1   Dipartimento di Ingegneria Dell'Informazione e Scienze Matematiche (DIISM), Università degli Studi di Siena, 53100 Siena, Italy
2   Department of Computer Science and Technology, University of Cambridge, Cambridge CB3 0FD, UK
3   Magnetic Resonance Center (CERM) and Department of Chemistry "Ugo Schiff", University of Florence, Sesto Fiorentino, 50019 Florence, Italy
4   Consorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine (C.I.R.M.M.P.), Sesto Fiorentino, 50019 Florence, Italy
*   Correspondence: luchinat@cerm.unifi.it (C.L.); pl219@cam.ac.uk (P.L.)
†   These authors contributed equally to this work.
‡   The PROPAG-AGEING Consortium is listed in acknowledgements.

Abstract: Biological aging can be affected by several factors such as drug treatments and pathological conditions. Metabolomics can help in the estimation of biological age by analyzing the differences between predicted and actual chronological age in different subjects. In this paper, we compared three different and well-known machine learning approaches—SVM, ElasticNet, and PLS—to build a model based on the ^{1}H-NMR metabolomic data of serum samples, able to predict chronological age in control individuals. Then, we tested these models in two pathological cohorts of de novo and advanced PD patients. The discrepancies observed between predicted and actual age in patients are interpreted as a sign of a (pathological) biological aging process.

Keywords: machine learning; metabolomics aging; spectrum; metabolites; lipids; Parkinson's disease; biological age

1. Introduction: Background and Objective of the Work

Biological age and its subsequent estimation are concepts that have found great interest in the bioinformatics literature over the last 50 years [1]. The dichotomy between "real chronological age" and the age expressed by the "true global state" [1] of the organisms has in fact been shown to be one of the most crucial biomarkers for possible pathological states. It is well known that chronological age represents one of the most important risk factors when it comes to predicting negative clinical outcomes [2]. At the same time, individuals of the same chronological age can present extremely different biological aging states, which can lead to completely different likelihood predictions when it comes to evaluating mortality and negative outcome risk. To generalize and to give a more comprehensive definition, biological age can be defined as the age resulting from a prediction produced by a statistical model incorporating a set of age-dependent variables and biomarkers. In recent years, several studies have explored the dichotomy between biological age and chronological age, focusing on different types of omics and imaging biomarkers. To give an example, Cole et al. [3] estimated biological age using brain MRI scans. In the paper, they introduced the concept of brain aging which represents a crucial step to investigate specific pathologies and diseases that might affect brain structure and cognition. A further notable example in the literature is the seminal Horvath clock, built using methylation information [4]. This sets the ground for the new concept of biological age estimation from molecular data.

Since then, other estimators based on epigenetics and DNA methylation have been developed such as PhenoAGE [5] and GrimAge [6]. These models are currently representing a relevant field of research in the area of epigenetics and aging. Analogously, metabolomics is currently seeing an increasing trend of interest in the development and study of models that could predict biological age starting from serum or plasma metabolomic data. The human metabolome has been in fact proven to correlate with (and possibly to predict) several phenotypical factors, such as gender [7,8], BMI [9], and age-related pathological conditions, for example, Parkinson's [10,11] and Alzheimer's [12,13] diseases. Considering the interdependence between age and metabolism, some studies have concentrated on this research topic. For example, linear regression models were used [14] to grasp the relationship between age and the levels of a panel of metabolites, using also BMI as a covariate. Other similar approaches have been developed [15–19]. Due to the established correlations between age and metabolomic data, a few later studies have focused on the development of predictive models to estimate biological age. In ref. [2], the authors presented a study where urine metabolome data were used as a biological age estimator. The authors started with a healthy cohort and then tested the regression model in a sample of obese individuals undergoing bariatric surgery [2]. In [20], instead, the authors focused on the estimation of age from urine samples analyzed in a cohort of 301 healthy individuals. Here, we present an attempt to estimate the biological age using metabolomic data acquired on serum samples. A strength of our approach is the use of a collection of different cohorts of Parkinson's disease (PD) patients plus control individuals, recruited through the PROPAG-AGEING Horizon 2020 Project [21]. Previous results obtained with the PROPAG-AGEING cohorts have already identified promising biomarkers for the discrimination between controls and PD patients using miRNA [22], metabolites, and lipoproteins [23]. In this study, the aim was to develop a biological age prediction model to study the discrepancies observed between age estimated in control individuals and age predicted in patients, using a nuclear magnetic resonance (NMR)-based metabolomic approach. In the present work, we introduced several novelties. To the best of our knowledge, this is the first time that age estimation through metabolomics is used to evaluate the age of Parkinson's patients. The model was, as previously mentioned, estimated using a cohort of non-PD control subjects and, subsequently, applied to predict age in two cohorts of, respectively, de novo and advanced PD patients. To estimate whether (and to what extent) there is a discrepancy between actual and estimated chronological age in PD patients, we implemented three different types of machine learning techniques, on the one hand, and the well-known Klemera–Doubal age estimator [1], on the other hand. We decided to do this in light of possibly then integrating further omics, such as methylation, in our analysis, with the aim of estimating biological aging such as the Horvath clock biological age estimator. Moreover, as a further point of novelty, we used both the profile of identified metabolites and lipoproteins (profiling approach) as well as the whole NMR spectra (fingerprinting approach) as input features for the modeling. Indeed, to the best of our knowledge, no age predictors have been built so far using the entire NMR spectrum directly; thus, the promising performance obtained opens the way to the possibility of directly using an untargeted "omic" approach based on the NMR-metabolic fingerprint for age prediction, instead of relying on a selection of biomarkers [24].

2. Materials and Methods

2.1. Study Participants

In the present study, a total of 675 serum samples were collected from individuals belonging to 4 different PROPAG-AGEING cohorts, including healthy subjects (Hs), centenarians (Cent) and their offspring (CentOs), siblings of PD patients (only siblings with a prodromal probability score of less than 10% were included in the study), de novo drug-naive PD patients (dn^2PD), and advanced PD patients undergoing dopaminergic treatment (advPD). Detailed PROPAG-AGEING cohorts' descriptions, patients' recruitment, and di-

agnosis are reported in [21]. The distribution of the samples included in our study per cohorts are reported in Table 1.

Table 1. Design and characteristics of the cohorts included in the study.

Cohort	Tot	N°	F/Tot	Mean Age F (Max; Min)	Mean Age M (Max; Min)
Hs	118	UNIBO (39) UMG-GOE (79)	54/118	66.7 (82.5; 52)	68.2 (85; 49)
Cent	57	UNIBO (39)	39/57	105.2 (112.3; 100)	102.9 (106.3; 100)
CentOs	46	UNIBO (39)	29/46	70.7 (89; 55)	71.1 (84; 58)
Sib	199	AUSL-ISNB (93) SAS (106)	115/199	59.8 (90; 23)	59.2 (84; 23)
dn^{2}PD	233	UMG-GOE (228) SAS (5)	109/233	65.1 (84; 29)	64.8 (87; 39)
advPD	22	UMG-GOE (22)	7/22	66.7 (77; 52)	70.0 (84; 59)

Hs: healthy subjects; Cent: centenarians; CentOs: centenarians' offspring; Sib: siblings; dn^{2}PD: de novo drug-naive PD patients; advPD: advanced PD patients under dopaminergic treatment; UNIBO: Alma Mater Studiorum, Università di Bologna (IT); UMG-GOE: Universitaetsmedizin Goettingen, Georg-August-Universitaet Goettingen, Stiftung Oeffentlichen Rechts (DE); AUSL-ISNB: Azienda Unita' Sanitaria Locale Di Bologna (IT); SAS: Servicio Andaluz De Salud (ES); N°: number of subjects per cohort; F: females; M: males.

All PD patients involved in PROPAG-AGEING underwent deep phenotyping, including international standards of motor classification (Hoehn and Yahr stages) [25] and Unified Parkinson's Disease Rating Scale (UPDRS) scores [26], used to describe the disease severity. Demographics and distribution of disease severity among PD patients are reported in Table 2.

Table 2. Characteristics of PD patients and severity scores. Total UPDRS score includes various items contributing to four subscales: (I) mentation, behavior, and mood; (II) activities of daily living; (III) motor symptoms; and (IV) complications of therapy.

	dn^{2}PD		advPD		
	Mean	sd	Mean	sd	*p*-Value
age	65.24	10.09	68.95	7.33	0.11
Hoehn and Yahr Scale	1.51	1.07	3.13	0.60	4.52×10^{-10}
UPDRS I	1.63	1.76	5.05	3.10	2.74×10^{-10}
UPDRS II	6.83	5.80	19.45	6.66	1.98×10^{-15}
UPDRS III	17.42	13.92	34.41	15.77	1.11×10^{-5}
UPDRS IV	0.62	1.39	5.15	4.22	2.89×10^{-18}
UPDRS sum	25.57	20.28	62.10	22.08	2.75×10^{-11}
Duration of the disease (years)	RD	RD	9.32	2.78	/
BMI	27.18	4.79	25.95	3.72	0.28

RD: Recently diagnosed.

2.2. Ethical Issues

The study was conducted according to the Declaration of Helsinki and with informed written consent provided by all subjects. The study was approved by the ethics committee of the Physician's Board Hesse, Germany (Approval No. FF89/2008 for DeNoPa), the University Medical Center Goettingen, Germany (Approval No. 9/7/04 and 36/7/02 for the Kassel cohort), the ISNB (Italy) ethics committee (No. of approval 16018 of May 2016), and the SAS (Spain) ethical committee (No. of approval 2014/PI173 of September 2016).

2.3. Metabolomics Analyses

NMR sample preparation was performed following standard procedures [24,27,28]. NMR spectra were acquired for all samples using a Bruker 600 MHz (5 mm PATXI ^{1}H-^{13}C-^{15}N and ^{2}H-decoupling probe including a z-axis gradient coil) equipped with an automatic refrigerated sample changer (SampleJet, Bruker BioSpin, Billerica, MA, USA). Temperature

was equilibrated using a BTO 2000 thermocouple at 310 K keeping each serum sample for 5 min inside the probe before the measurement.

Each serum sample was analyzed using a NOESY 1D presat (noesygppr1d.comp; Bruker BioSpin) pulse sequence using 32 scans, 98,304 data points, a spectral width of 18,028 Hz, an acquisition time of 2.7 s, a relaxation delay of 4 s, and a mixing time of 0.01 s and suppressing the water peak at 4.702 ppm.

2.4. Data Processing

Free induction decays were multiplied by an exponential function that was equal to a 0.3 Hz line-broadening factor before the Fourier transform was applied. TopSpin 3.6.2 was used to automatically adjust transformed spectra for phase and baseline aberrations (Bruker BioSpin). Spectra were then calibrated to the anomeric glucose doubled (δH 5.24 ppm (parts per million)) and bucketed in the range between 10.0 ppm and 0.2 ppm into 0.02 ppm chemical shift bins (each one of 87 points) using AssureNMR (v 2.2) software (Bruker BioSpin). The residual water signal region in the spectra (from 4.68 ppm to 4.84 ppm) was removed, resulting in a final data matrix of bucketed spectra of 479 columns.

2.5. Serum Metabolite and Lipoprotein Quantification

The AVANCE Bruker IVDr (Clinical Screening and In Vitro Diagnostics research, Bruker BioSpin) platform was used to measure the fractions and subfractions of 112 lipoproteins and twenty-eight metabolites [29,30]. Different lipoproteins (VLDL, LDL, IDL, and HDL) and lipoprotein subclasses (VLDL-1 to VLDL-5, LDL-1 to LDL-6, and HDL-1 to HDL-4) with a total of 15 subclasses (VLDL-1 to VLDL-5, LDL-1 to LDL-6, and HDL-1 to HDL-4) were taken into consideration for all blood samples. The provided results include the concentrations of lipids (total cholesterol, free cholesterol, phospholipids, and triglycerides) found in each fraction for each major class and subclass. While Apo-B concentrations were computed for VLDL, IDL classes, and all LDL subclasses, ApoA1 and ApoA2 concentrations were estimated for the HDL class and each related subclass.

2.6. Age Prediction Using Machine Learning Models

The data used for the present analysis consisted of the matrix of the bucketed 1D NMR spectra (675 × 479) and the matrix of the quantified metabolites and lipoproteins (675 × 140) in serum samples, as described in the previous sections (Sections 2.4 and 2.5).

The metabolite and lipoprotein matrix as well as the spectra matrix were normalized using the Z-score transformation (subtracting the mean and scaling by the standard deviation).

To perform data analysis, we used three well-known and validated machine learning methodologies: support vector machine (SVM) with linear kernel, generalized linear models (glmnet with ElasticNet regression), and PLS. SVM was implemented with a linear kernel using the scikit library in python version 3.1. ElasticNet and PLS models were implemented using R version 3.2.1. with glmnet package version 4.1.1 and PLS package version 2.7.3. Hyperparameters for SVM and ElasticNet were optimized using cross-validation. The number of PLS components was set to 6 after a few preliminary trials. The Klemera and Doubal method was implemented using the R function "TrueTrait" included in the package WGCNA (R package version 1.69).

To evaluate the algorithms, cross-validation was employed. Root mean squared error (RMSE) and R^2 for all the methods were reported.

To build the machine learning models, we trained the algorithms using the controls (Ctr, non-PD subjects) of the PROPAG-AGEING cohort, both with the dataset of binned spectra and with the dataset of metabolite/lipoprotein input features. The control cohort that the training set was composed of included healthy controls (Hs), siblings with risk scores of less than 10% (Sib), centenarians (Cent), and centenarians' offspring (CentOs). The distribution of individuals per cohort is reported in Table 1. In particular, the training set is composed of 420 subjects. Each model trained using this dataset was subsequently tested in the dn^{2}PD and advPD cohorts, which were used as test sets. The performance of

the models for chronological age prediction was assessed in both controls (training) and patients (test). The rationale was to use the discrepancy between the actual chronological age and the predicted chronological age as a proxy to estimate the health status of the diseased individuals, in the hypothesis that PD patients are (from a metabolic point of view) more like aged controls than younger individuals (i.e., they may have a higher biological than chronological age).

The overall representation of the machine learning pipeline is reported in Figure 1.
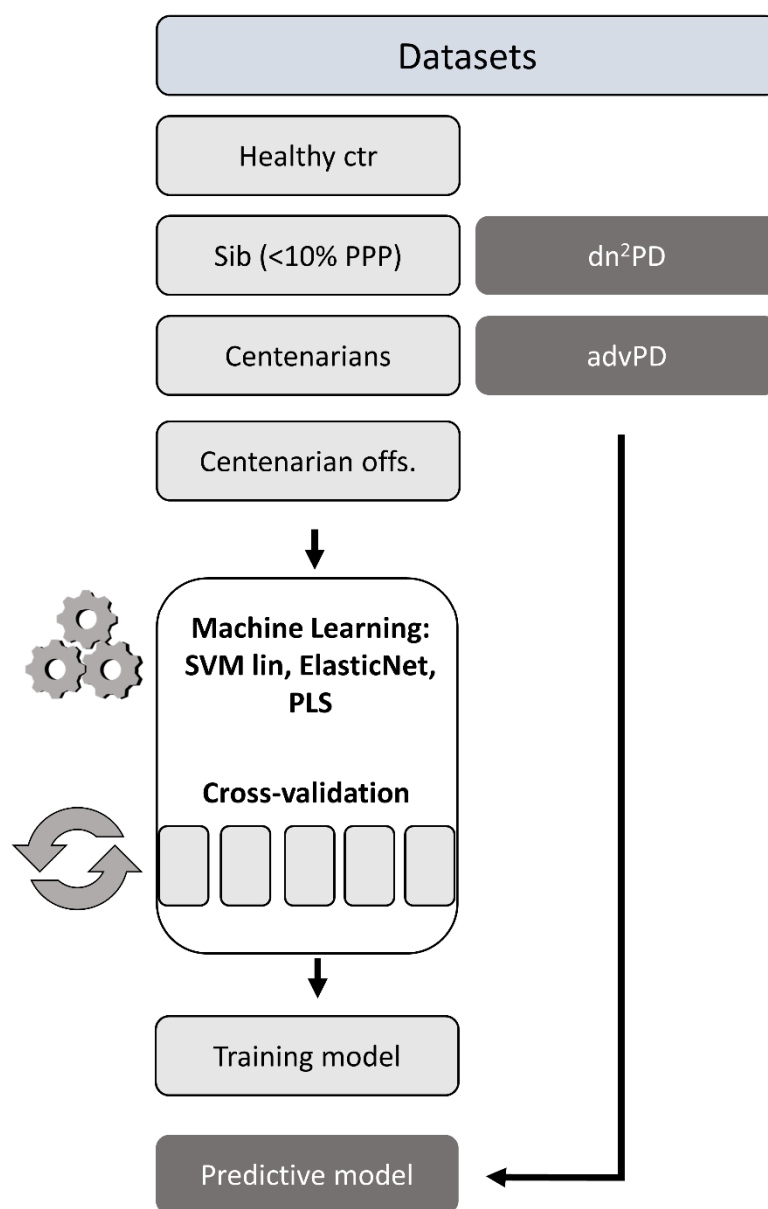
Figure 1. Schematic representation of the age prediction models.

3. Results

3.1. Age Prediction Using Fingerprints and Profiles

Tables 3 and 4 report R^2 (Supplementary Figures S1–S6) and RMSE for the binned spectra and for the metabolite/lipoprotein matrices used as input datasets. The tables show the best-performing models trained and tested in the control population for each of the algorithms considered. Subsequently, we reported the performances of these models when tested on the patient population (using dn^{2}PD patients and advPD patients) to evaluate whether the predicted age is consistent or not with the actual age and to enlighten a possible

mismatch due to the disease (i.e., a biological age not corresponding to the chronological one). The best-performing model, trained on the control population, was used and tested in the patient population. For what concerns the models based on the bucket matrix, the best-performing methods in the healthy dataset are SVM and PLS, with an R^2 of 0.865 and 0.825, respectively. The lowest RMSE (6.273) was obtained for SVM, which can therefore be defined as the best-performing model in the control dataset. This score becomes much larger when dn^2PD patients and advPD patients are used as a test set. This means that a high discrepancy between predicted and actual chronological age emerges in patients affected by PD.

Table 3. Table reporting performance for SVM, PLS, and ElasticNet. The Klemera–Doubal baseline is also presented. Binned spectra are used as input features.

Model Based on Spectrum	Ctr		dn^2PD		advPD	
	R^2	RMSE	R^2	RMSE	R^2	RMSE
SVM Linear	0.865	6.273	0.208	11.209	0.037	13.057
ElasticNet	0.811	7.466	0.255	12.488	0.049	12.789
PLS	0.825	7.126	0.219	12.963	0.129	10.348
Klemera–Doubal y.true1	0.161	38.929	0.035	25.84	0.004	25.591
Klemera–Doubal y.true2	0.301	25.936	0.036	30.155	0.0001	29.861

Table 4. Table reporting performance for SVM, PLS, and ElasticNet together with the baseline model of Klemera–Doubal, using metabolites/lipoproteins as input features.

Model Based on Metabolites	Ctr		dn^2PD		advPD	
	R^2	RMSE	R^2	RMSE	R^2	RMSE
SVM Linear	0.735	8.76	0.314	12.651	0.138	10.961
ElasticNet	0.756	8.422	0.236	11.044	0.014	13.562
PLS	0.739	8.704	0.095	15.157	0.043	10.423
Klemera–Doubal y.true1	0.046	77.305	0.001	110.239	0.007	111.439
Klemera–Doubal y.true2	0.318	24.903	0.039	27.304	0.0002	31.188

We also report the estimates obtained with the Klemera–Doubal model, showing how all three machine learning approaches (SVM, PLS, and ElasticNet) significantly outperform this model in terms of R^2.

Similar considerations can be applied when using as input features the matrix of metabolites/lipoproteins, as shown in Table 4.

For what concerns metabolite/lipoprotein predictions, as shown in Table 4, performances are slightly worse than the models built on binned spectra, reaching a maximum R^2 of 0.765 for the ElasticNet method with the lowest RMSE in the Ctr model. Again, dn^2PD and advPD show a higher discrepancy when the chronological age is predicted.

Therefore, we used the best SVM output model to identify those subjects with a predicted age that is at least 6 years over/underestimated by the model with respect to the chronological age, resulting in a total of 264 subjects.

Of them, 7 advPD patients are predicted as older (with a mean difference between real and predicted age of 15.45 years and a standard deviation (SD) of 10.5) accounting for 31.81% of the total advPD population, while 9 advPD are predicted as younger (11.4 years; 4.87 years) accounting for 40.97% of the advPD population; 95 dn^2PD patients are predicted as older (14.7 years; 6.4 years) accounting for 40.775% of the dn^2PD population, while 39 dn^2PD patients are predicted as younger with respect to their chronological age (10.0 years; 3.2 years) accounting for 16.73% of the total dn^2PD population; 58 of the control population are predicted as older (10.5 years; 4.8 years), and 56 of the Ctr are predicted as younger (10.4 years; 9.3 years) accounting, respectively, for 13.8% and 13.3% of the total Ctr population.

These data suggest that the percentage of subjects misclassified as older is about 18–27% higher in patients than in the control group.

A schematic representation of the above data is reported in Table 5.

Table 5. Difference between the chronological age and the predicted age (years) for each study group. Overestimated are the individuals that are predicted at least 6 years older than their actual chronological age; underestimated are individuals that are predicted at least 6 years younger than their actual chronological age. Percentage reported refers to the total of misclassified individuals per group.

	Overestimated				Underestimated			
	%Subj./Group	RA(m ± SD)	PA(m ± SD)	PA-RA	%Subj./Group	RA(m ± SD)	PA(m ± SD)	RA-PA
advPD	31.8	64.29 ± 7.54	79.74 ± 15.49	15.45 ± 10.51	40.9	72.33 ± 6.58	60.94 ± 5.09	11.39 ± 4.87
dn²PD	40.8	55.84 ± 10.81	70.54 ± 9.74	14.7 ± 6.40	16.7	74.23 ± 5.59	64.2 ± 6.30	10.01 ± 3.24
Ctr	13.8	55.81 ± 15.78	66.36 ± 16.17	10.55 ± 4.84	13.3	84.86 ± 16.45	74.41 ± 16.2	10.44 ± 3.88

RA: real age in years; m: mean; SD: standard deviation; PA: predicted age in years.

3.2. Correlation between Predicted Ages and Disease Severity

Correlations among severity scores, predicted and actual chronological age, BMI, and sex were calculated.

In Figure 2, the resulting correlation matrix is reported, considering as input both the binned spectra and the metabolite/lipoprotein data acquired from the cohorts of advPD patients. As we can see from the plot, the correlation with severity scores becomes much higher when the predicted age is considered with respect to the chronological age.
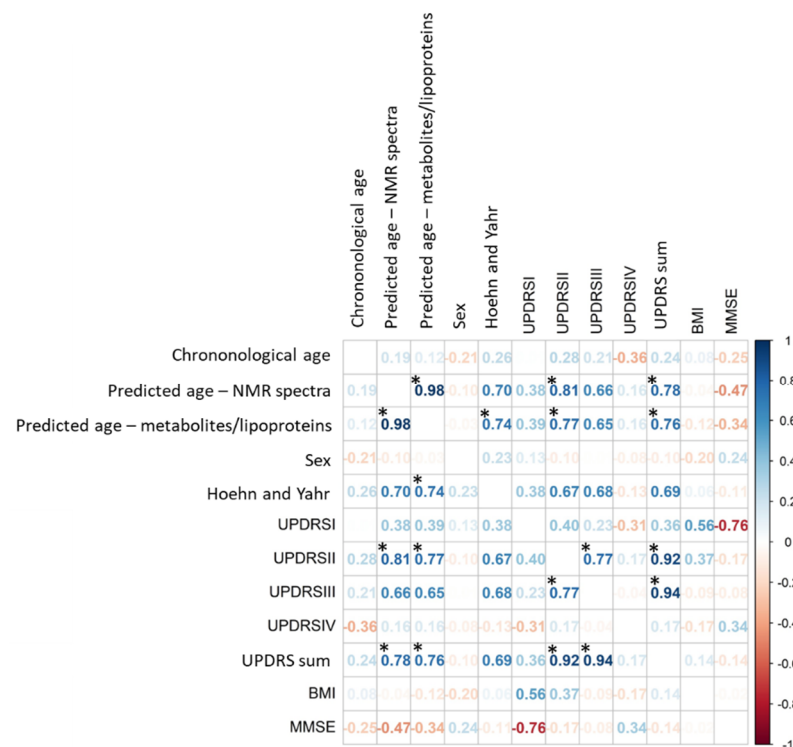
Figure 2. Pearson's correlation coefficients among predicted age (estimated using PLS), severity scores, and other demographic and physiological variables. Data are obtained using the advPD cohort. Asterisks represent p-values < 0.05.

As we can see from Figure 2, the correlation of severity scores with predicted (biological) age is much higher than its correlation with actual chronological age. We therefore attempted the same correlation plot (Figure 3), considering the dn²PD patients and their respective severity scores. In this case, correlations are not as strong as in the case of

the correlations in advPD patients. This can be attributed to the fact that dn^2PD patients are at the beginning of the disease and, as can be seen from Table 2, they are characterized by very low severity scores compared to advPD patients.
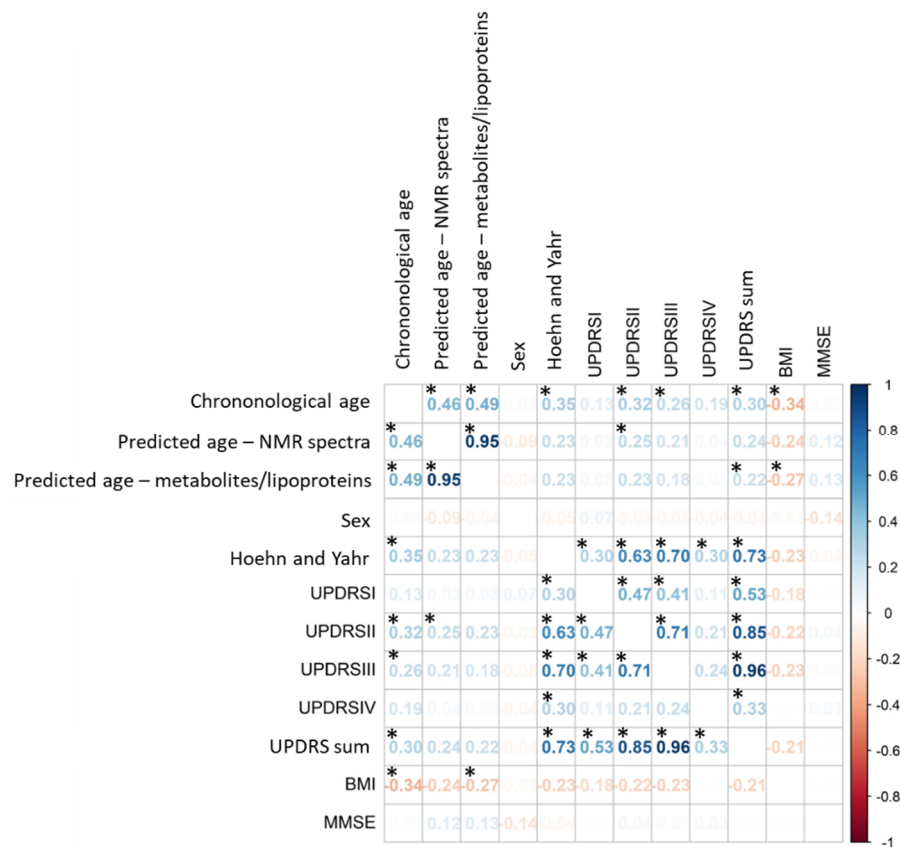

Figure 3. Pearson's correlation coefficients among predicted age (estimated using PLS), severity scores, and other demographic and physiological variables. Data are obtained using the dn^2PD cohort. Asterisks represent *p*-values < 0.05.

4. Discussion and Limitations

Biological aging can be defined as a collection of interconnected molecular and cellular changes associated with aging, and it is thought to induce physiological degeneration, which is at the root of a variety of age-related health disorders, according to the emerging area of geroscience. Indeed, while some people get age-related diseases such as Parkinson's disease, others, such as centenarians, can live to 100 years old and beyond in good health. There is a spectrum of intermediate phenotypes between these two extremes, including those with more or less pronounced subclinical signs of disorders [21]. In this study, we decided to evaluate the biological age of a large population of healthy subjects, centenarians and their offspring, siblings of PD patients, and two groups of PD patients including de novo drug-naive PD patients and advanced PD patients under dopaminergic treatment. To our knowledge, this is the first time that a metabolomic-based age estimation has been utilized to assess the age of Parkinson's disease patients. The model was developed using data from a collection of healthy non-PD cohorts and then tested using data from the de novo drug-naive and advanced PD patients to assess how large is the difference between actual and predicted age in PD patients and in healthy subjects. Four algorithms were used for the prediction of biological aging: ElasticNet, PLS, SVM, and Klemera–Doubal. SVM resulted to be the best-performing one, especially when coupled with the NMR fingerprint approach. Interestingly, we found that the predictive model built on healthy subjects shows a quite good agreement between the actual and the predicted chronological age. Conversely, when we attempted to predict the age of PD patients, the model showed a larger discrep-

ancy. Moreover, the error is larger in dn^{2}PD patients than in advPD patients. Reasonably, this could be due to the presence of the dopaminergic treatment in the advPD group, which interferes with and/or reduces the gap between chronological and predicted age.

Further, it seems that working with a larger data matrix, such as the full NMR spectra, works better than working using only some components of the spectra. The full NMR spectrum is considered the "fingerprint" of all (assigned or unassigned) detectable metabolites and lipoproteins present in that biological sample. Indeed, the fingerprinting approach is essentially utilized in metabolomics to provide sample classification. In contrast, the determination of the concentrations of all quantifiable metabolites in a biological sample, defined as "profiling", is generally used to provide information regarding the metabolic pathways involved in specific pathological or physiological conditions [24]. However, the molecules that can be quantified by profiling are far fewer than those that contribute to the fingerprint. Therefore, it was not completely unexpected that the fingerprinting approach performed better in sample classification. Interestingly, we found a correlation between the predicted chronological age and the severity scores (Hoehn and Yahr scale and UPDRS) in advPD patients. This correlation is stronger than that calculated from the actual chronological age. This evidence suggests a possible use of metabolomics in the determination of a biological aging process that is also descriptive of the progress of PD. However, one of the main limitations of the study is the limited number of advPD patients. Further, linking the metabolomic-based age estimation to the altered metabolic pathways could improve our understanding of the aging process in healthy people and patients with cognitive disorders. A possible future extension of this work could involve the use of larger cohorts of controls and advanced-disease patients and include other cognitive disorders such as Alzheimer's and dementia.

References

1. Klemera, P.; Doubal, S. A New Approach to the Concept and Computation of Biological Age. *Mech. Ageing Dev.* **2006**, *127*, 240–248. [CrossRef] [PubMed]
2. Hertel, J.; Friedrich, N.; Wittfeld, K.; Pietzner, M.; Budde, K.; Van der Auwera, S.; Lohmann, T.; Teumer, A.; Völzke, H.; Nauck, M.; et al. Measuring Biological Age via Metabonomics: The Metabolic Age Score. *J. Proteome Res.* **2016**, *15*, 400–410. [CrossRef] [PubMed]
3. Cole, J.H.; Franke, K. Predicting Age Using Neuroimaging: Innovative Brain Ageing Biomarkers. *Trends Neurosci.* **2017**, *40*, 681–690. [CrossRef] [PubMed]
4. Horvath, S.; Raj, K. DNA Methylation-Based Biomarkers and the Epigenetic Clock Theory of Ageing. *Nat. Rev. Genet.* **2018**, *19*, 371–384. [CrossRef] [PubMed]
5. Levine, M.E.; Lu, A.T.; Quach, A.; Chen, B.H.; Assimes, T.L.; Bandinelli, S.; Hou, L.; Baccarelli, A.A.; Stewart, J.D.; Li, Y.; et al. An Epigenetic Biomarker of Aging for Lifespan and Healthspan. *Aging (Albany NY)* **2018**, *10*, 573–591. [CrossRef]
6. Lu, A.T.; Quach, A.; Wilson, J.G.; Reiner, A.P.; Aviv, A.; Raj, K.; Hou, L.; Baccarelli, A.A.; Li, Y.; Stewart, J.D.; et al. DNA Methylation GrimAge Strongly Predicts Lifespan and Healthspan. *Aging (Albany NY)* **2019**, *11*, 303–327. [CrossRef]
7. Auro, K.; Joensuu, A.; Fischer, K.; Kettunen, J.; Salo, P.; Mattsson, H.; Niironen, M.; Kaprio, J.; Eriksson, J.G.; Lehtimäki, T.; et al. A Metabolic View on Menopause and Ageing. *Nat. Commun.* **2014**, *5*, ncomms5708. [CrossRef]
8. Di Cesare, F.; Tenori, L.; Meoni, G.; Gori, A.M.; Marcucci, R.; Giusti, B.; Saccenti, E. Lipid and Metabolite Correlation Networks Specific to Clinical and Biochemical Covariate Show Differences Associated with Sexual Dimorphism in a Cohort of Nonagenarians. *GeroScience* **2022**, *44*, 1109–1128. [CrossRef]
9. Xie, B.; Waters, M.J.; Schirra, H.J. Investigating Potential Mechanisms of Obesity by Metabolomics. *J. Biomed. Biotechnol.* **2012**, *2012*, 805683. [CrossRef]
10. Scholefield, M.; Unwin, R.D.; Cooper, G.J.S. Shared Perturbations in the Metallome and Metabolome of Alzheimer's, Parkinson's, Huntington's, and Dementia with Lewy Bodies: A Systematic Review. *Ageing Res. Rev.* **2020**, *63*, 101152. [CrossRef]
11. Hu, L.; Dong, M.-X.; Huang, Y.-L.; Lu, C.-Q.; Qian, Q.; Zhang, C.-C.; Xu, X.-M.; Liu, Y.; Chen, G.-H.; Wei, Y.-D. Integrated Metabolomics and Proteomics Analysis Reveals Plasma Lipid Metabolic Disturbance in Patients With Parkinson's Disease. *Front. Mol. Neurosci.* **2020**, *13*, 80. [CrossRef] [PubMed]
12. Huo, Z.; Yu, L.; Yang, J.; Zhu, Y.; Bennett, D.A.; Zhao, J. Brain and Blood Metabolome for Alzheimer's Dementia: Findings from a Targeted Metabolomics Analysis. *Neurobiol. Aging* **2020**, *86*, 123–133. [CrossRef] [PubMed]
13. Vignoli, A.; Paciotti, S.; Tenori, L.; Eusebi, P.; Biscetti, L.; Chiasserini, D.; Scheltens, P.; Turano, P.; Teunissen, C.; Luchinat, C.; et al. Fingerprinting Alzheimer's Disease by 1H Nuclear Magnetic Resonance Spectroscopy of Cerebrospinal Fluid. *J. Proteome Res.* **2020**, *19*, 1696–1705. [CrossRef] [PubMed]
14. Yu, Z.; Zhai, G.; Singmann, P.; He, Y.; Xu, T.; Prehn, C.; Römisch-Margl, W.; Lattka, E.; Gieger, C.; Soranzo, N.; et al. Human Serum Metabolic Profiles Are Age Dependent. *Aging Cell* **2012**, *11*, 960–967. [CrossRef]
15. Ishikawa, M.; Maekawa, K.; Saito, K.; Senoo, Y.; Urata, M.; Murayama, M.; Tajima, Y.; Kumagai, Y.; Saito, Y. Plasma and Serum Lipidomics of Healthy White Adults Shows Characteristic Profiles by Subjects' Gender and Age. *PLoS ONE* **2014**, *9*, e91806. [CrossRef]
16. Lawton, K.A.; Berger, A.; Mitchell, M.; Milgram, K.E.; Evans, A.M.; Guo, L.; Hanson, R.W.; Kalhan, S.C.; Ryals, J.A.; Milburn, M.V. Analysis of the Adult Human Plasma Metabolome. *Pharmacogenomics* **2008**, *9*, 383–397. [CrossRef]

17. Menni, C.; Kastenmüller, G.; Petersen, A.K.; Bell, J.T.; Psatha, M.; Tsai, P.-C.; Gieger, C.; Schulz, H.; Erte, I.; John, S.; et al. Metabolomic Markers Reveal Novel Pathways of Ageing and Early Development in Human Populations. *Int. J. Epidemiol.* **2013**, *42*, 1111–1119. [CrossRef]

18. Collino, S.; Montoliu, I.; Martin, F.-P.J.; Scherer, M.; Mari, D.; Salvioli, S.; Bucci, L.; Ostan, R.; Monti, D.; Biagi, E.; et al. Metabolic Signatures of Extreme Longevity in Northern Italian Centenarians Reveal a Complex Remodeling of Lipids, Amino Acids, and Gut Microbiota Metabolism. *PLoS ONE* **2013**, *8*, e56564. [CrossRef]

19. Swann, J.R.; Spagou, K.; Lewis, M.; Nicholson, J.K.; Glei, D.A.; Seeman, T.E.; Coe, C.L.; Goldman, N.; Ryff, C.D.; Weinstein, M.; et al. Microbial-Mammalian Cometabolites Dominate the Age-Associated Urinary Metabolic Phenotype in Taiwanese and American Populations. *J. Proteome Res.* **2013**, *12*, 3166–3180. [CrossRef]

20. Rist, M.J.; Roth, A.; Frommherz, L.; Weinert, C.H.; Krüger, R.; Merz, B.; Bunzel, D.; Mack, C.; Egert, B.; Bub, A.; et al. Metabolite Patterns Predicting Sex and Age in Participants of the Karlsruhe Metabolomics and Nutrition (KarMeN) Study. *PLoS ONE* **2017**, *12*, e0183228. [CrossRef]

21. Pirazzini, C.; Azevedo, T.; Baldelli, L.; Bartoletti-Stella, A.; Calandra-Buonaura, G.; Dal Molin, A.; Dimitri, G.M.; Doykov, I.; Gómez-Garre, P.; Hägg, S.; et al. A Geroscience Approach for Parkinson's Disease: Conceptual Framework and Design of PROPAG-AGEING Project. *Mech. Ageing Dev.* **2021**, *194*, 111426. [CrossRef] [PubMed]

22. Zago, E.; Dal Molin, A.; Dimitri, G.M.; Xumerle, L.; Pirazzini, C.; Bacalini, M.G.; Maturo, M.G.; Azevedo, T.; Spasov, S.; Gómez-Garre, P.; et al. Early Downregulation of Hsa-MiR-144-3p in Serum from Drug-Naïve Parkinson's Disease Patients. *Sci. Rep.* **2022**, *12*, 1330. [CrossRef]

23. Meoni, G.; Tenori, L.; Schade, S.; Licari, C.; Pirazzini, C.; Bacalini, M.G.; Garagnani, P.; Turano, P.; PROPAG-AGEING Consortium; Trenkwalder, C.; et al. Metabolite and Lipoprotein Profiles Reveal Sex-Related Oxidative Stress Imbalance in de Novo Drug-Naive Parkinson's Disease Patients. *NPJ Parkinsons Dis.* **2022**, *8*, 14. [CrossRef] [PubMed]

24. Vignoli, A.; Ghini, V.; Meoni, G.; Licari, C.; Takis, P.G.; Tenori, L.; Turano, P.; Luchinat, C. High-Throughput Metabolomics by 1D NMR. *Angew. Chem.-Int. Edit.* **2019**, *58*, 968–994. [CrossRef] [PubMed]

25. Hoehn, M.M.; Yahr, M.D. Parkinsonism: Onset, Progression and Mortality. *Neurology* **1967**, *17*, 427–442. [CrossRef] [PubMed]

26. Ebersbach, G.; Baas, H.; Csoti, I.; Müngersdorf, M.; Deuschl, G. Scales in Parkinson's Disease. *J. Neurol.* **2006**, *253* (Suppl. S4), IV32–IV35. [CrossRef]

27. Ghini, V. NMR for sample quality assessment in metabolomics. *New Biotechnol.* **2019**, *52*, 34. [CrossRef]

28. Takis, P.G.; Ghini, V.; Tenori, L.; Turano, P.; Luchinat, C. Uniqueness of the NMR Approach to Metabolomics. *TrAC Trends Anal. Chem.* **2019**, *120*, 115300. [CrossRef]

29. Reproducible Metabolite Quantification in Plasma/Serum. Available online: https://www.bruker.com/products/mr/nmr-preclinical-screening/biquant-ps.html (accessed on 2 May 2019).

30. Lipoprotein Subclass Analysis Enabling Tools on the IVDr Platform. Available online: https://www.bruker.com/products/mr/nmr-preclinical-screening/lipoprotein-subclass-analysis.html (accessed on 2 May 2019).