Intelligent Systems with Applications 27 (2025) 200571

Contents lists available at ScienceDirect

Intelligent Systems with Applications

journal homepage: www.journals.elsevier.com/intelligent-systems-with-applications

Application of LSTM and GRU neural networks to improve peristaltic pump dosing accuracy

Davide Privitera [a,*], Stefano Bellissima [b], Sandro Bartolini [a]

[a] Department of Information Engineering and Mathematics, University of Siena, Via Roma 56, Siena, 53100, Italy
[b] Pharma Integration, Strada del Petriccio e Belriguardo 35, Siena, 53100, Italy

ARTICLE INFO

Keywords:
Peristaltic pump
Dosing accuracy
LSTM
GRU
Adaptive control

ABSTRACT

Peristaltic pumps (PP), widely acknowledged for their benefits in pharmaceutical contexts, face challenges in achieving optimal dosing accuracy. This investigation contributes novel insights for the improvement of dosing precision, identifying how to apply AI models, specifically focusing on Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) neural networks over a realistic span of target volumes. To provide a more accurate representation of real-world performance, we consider a modified root mean square error metric ($RMSE_{PP}$) that directly compares dispensed volumes to target volumes. Based on this the study delves into two main methodologies: an iterative retraining method, called Online Training, and Pre-trained approach. Online Training shows best results, especially for volumes below 1.0 ml, achieving 38.4% improvement in $RMSE_{PP}$ and 31.6% in standard deviation (STD). Pre-trained models are faster and exhibit promising outcomes especially for volumes above 1.0 ml, with a three-features approach delivering the best performance (13.8% and 4.6% improvements in $RMSE_{PP}$ and STD, respectively). Overall, the findings highlight the effectiveness of iterative learning techniques, particularly for smaller dosage amounts, which complements the good performance of non-AI approaches for larger ones.

1. Introduction

Peristaltic filling technology has gained significant popularity in recent decades, particularly within the pharmaceutical sector, due to its ability to handle sensitive bio-pharmaceutical products with minimal risk of cross-contamination (Lambert & Joergensen, 2008). In peristaltic pumps (PP), the product only contacts a single, easily replaceable section of tubing, which can be replaced post-use. This tubing, typically composed of elastomers or thermoplastics, is encased in a circular seat on the pump body and is compressed by the rotational movement of two or more rollers, as depicted in Fig. 1. Traditional dosing devices like volumetric pumps create contamination risks due to direct product contact with difficult-to-clean mechanical parts. FDA recall data from 2012–2023 highlights the criticality of these issues, with sterility-related problems accounting for 36.7% of all recalls, the highest single category (Patel et al., 2024). Among contamination-specific recalls, microbial contamination dominated (40.07%), with a combined 15% attributed to cross-contamination issues and presence of foreign substances, affecting various processes and equipment. High-profile cases further illustrate these risks, such as the New England Compounding Center incident that resulted in 751 fungal meningitis cases and 64 deaths from contaminated injectables (Natof & Pellegrini, 2025). The unique design of PP significantly mitigates these contamination concerns, as the product only contacts the sterile, disposable tubing.

This design, along with the ability to manage a wide range of filling volumes (0.1 to 250 ml) by adjusting tube size, makes PP an attractive choice for dispensing bio-pharmaceutical injectable drugs (Klespitz & Kovács, 2014). However, a significant concern with PP is their fill accuracy, particularly at low volumes (Gasoto et al., 2022; Klespitz & Kovács, 2014; Shieu et al., 2016).

Improving PP accuracy is crucial from both economic and patient safety perspectives. Economically, enhanced precision can significantly reduce product waste, a critical factor given the high costs associated with bio-pharmaceutical development and manufacturing. Farid et al. (2020) reported that ensuring a market success with a 12% clinical success rate (Phase I to approval) requires allocating approximately $60 million for pre-clinical to Phase II material preparation and $70 million for Phase III to regulatory review. Improved pump accuracy can lead to substantial savings by minimizing filling errors and reducing batch rejections, especially considering that process development and manufacturing costs represent 13%–17% of the R&D budget from pre-clinical trials to approval (Farid et al., 2020). The economic implications

* Corresponding author.
E-mail address: privitera@diism.unisi.it (D. Privitera).

https://doi.org/10.1016/j.iswa.2025.200571
Received 25 September 2024; Received in revised form 20 June 2025; Accepted 12 August 2025
Available online 20 August 2025
2667-3053/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

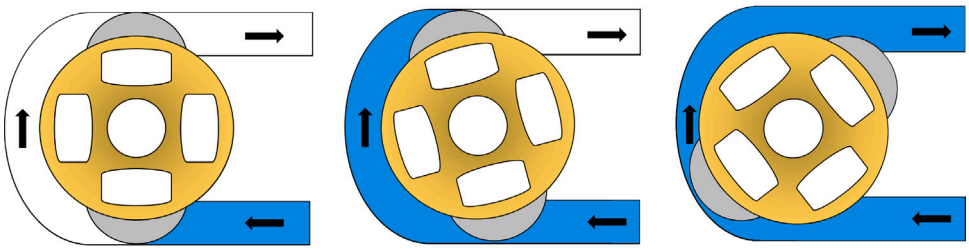

Fig. 1. Peristaltic pump working principle. The liquid flows through sterile tubing (blue) that is compressed by rotating rollers (gray). The sequential compression creates a positive displacement pumping action: (left) initial compression, (center) fluid progression, (right) complete cycle. The liquid only contacts the disposable tubing, eliminating contamination risks. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of accuracy improvements are particularly substantial also for high-value biopharmaceuticals. For example, with Zolgensma priced at $1.9 million per dose, even a modest 1% improvement in manufacturing precision would save approximately $19,000 per dose. At 800 doses annually, this translates to potential cost reductions of $15.2 million yearly (Nuijten, 2022). Similarly, for Patisiran (€8529.41 per 10 mg vial (Cozzio et al., 2023)) with annual production of 50,000 vials, a 1% accuracy improvement could reduce production costs by €4.26 million yearly while simultaneously increasing production capacity without additional capital investment. From a patient safety perspective, accurate dosing is essential to prevent medication errors and ensure proper therapeutic effects. The FDA guidance (U.S. Food and Drug Administration, 2015) highlights that inappropriate excess volume in injectable vials can lead to medication errors and misuse of leftover drug product. This is particularly critical for single-dose vials containing significantly more drug than required for a single dose, potentially resulting in underdosing or overdosing (Tariq et al., 2023). Furthermore, the need to combine several single-dose vials for a single patient dose may lead to medication errors and microbial contamination (U.S. Food and Drug Administration, 2015). The consequences can be severe, especially for drugs with narrow therapeutic indices.

In literature, while mechanical improvements for PP have been explored, there has been a notable gap in research applying advanced software-based approaches to enhance accuracy in industrial precision filling systems. By contrast, other manufacturing fields have successfully implemented Forecasting Compensatory Control (FCC) approaches, which compensate for errors through software rather than investing significant resources in mechanical improvements. These methods address both repetitive, systematic errors and predict non-repetitive stochastic variations, enabling correction of correlated dynamic inaccuracies. Notable applications include end milling (Wu & Ni, 1989), where laser sensors evaluate workpiece flatness with real-time tool position adjustments, wall thickness control (Wang et al., 2018), integrated circuit production (Yan & Huang, 2009), precision multi-axis motion systems (Hu et al., 2020), and thermal distortion compensation in additive manufacturing (Francis & Bian, 2019). Despite the demonstrated success of these techniques in various manufacturing domains, traditional PP accuracy improvements have historically focused on mechanical advancements: more precise roller mechanisms, better tubing materials, and sophisticated designs. While yielding some progress, these approaches often involve high costs with limited effectiveness, especially across diverse pharmaceutical production requirements (Ferretti et al., 2023; Gasoto et al., 2022; Klespitz & Kovács, 2014). To address this gap, previous work introduced an Adaptive Dosing Control System (ADCS) based on the ARIMA statistical model (Privitera et al., 2023). While ADCS shares the conceptual goal of error reduction through software-based compensation with FCC approaches, it introduces two fundamental distinctions: first, unlike traditional FCC methods that usually require detailed mathematical modeling of the system's internal dynamics, ADCS employs a black-box approach that focuses on input–output relationships, eliminating the need for complex system-specific formulations. Second, ADCS utilizes an adaptive approach that allows runtime adjustment of working parameters, continuously adapting to system variations without requiring a single fixed model to be applied uniformly across all conditions. This system operates through a closed-loop control mechanism that (i) measures actual dispensed volumes, (ii) predicts future dispensing behavior based on historical patterns without requiring knowledge of the underlying physical processes, and (iii) applies real-time compensations to adjust upcoming dispenses. The ARIMA-based implementation achieved up to 30% accuracy improvement for 1.2 ml volumes.

This work goes beyond statistical models, investigating a novel approach using Gated Recurrent Unit (GRU) (Cho et al., 2014) and Long Short-Term Memory (LSTM) (Hochreiter & Schmidhuber, 1997) models. These AI methods aim to better capture and exploit the underlying complex behavior of the dosing system, which may not be fully in the reach of conventional approaches featuring simpler underlying mathematical models and only a few parameters. By leveraging these techniques, we seek to improve the quality of compensation and, consequently, enhance the overall accuracy of filling systems in industrial applications.

In recent years, AI methods have gained widespread adoption across various fields due to their effectiveness and versatility in addressing complex modeling challenges. Techniques such as artificial neural networks, support vector machines, and deep learning algorithms have demonstrated remarkable success in predictive modeling tasks. For example, Cao et al. (2023) and Chen et al. (2021) discuss applications of Radial Basis Function Neural Networks for system prediction, while Francis and Bian (2019) explores AI approaches for thermal distortion prediction. Recent research trends further demonstrate the growing relevance of neural network architectures across diverse control applications characterized by complex dynamics and uncertainties, as exemplified in works by Banu et al. (2025) and Rajchakit et al. (2025), who explored applications for handling complex system behaviors under challenging conditions, like parameter uncertainties and external disturbances, enhancing precision and reliability.

The selection of LSTM and GRU models for this study is based on several key considerations. Firstly, the inherent nature of the problem involves sequential data, where each dosage is influenced by previous dosages. LSTM and GRU architectures are specifically designed to handle such time series data, excelling at capturing both short-term and long-term dependencies in sequences (Hochreiter & Schmidhuber, 1997). This capability encompasses complex underlying behaviors in the data. Their gating mechanisms allow them to selectively remember or forget information over sequences of varying lengths, making them ideal for modeling the intricate temporal dynamics of dosing systems. As Greff et al. (2017) demonstrate, LSTM and its variants consistently perform well across a wide range of sequence modeling tasks. Recent works have further demonstrated the continued relevance and efficacy of these models across diverse domains, including industrial applications. In the renewable energy sector, a study on wind turbine variable prediction (Buestán-Andrade et al., 2023) showed that the GRU model outperformed other architectures in predicting time series data from wind turbines. In power systems, Wu et al. (2024) found that LSTM-based models achieved high detection accuracy for

fault detection. In text summarization, Bayat and Işık (2023) found that the GRU model exhibited a better balance between accuracy and conciseness. In oceanographic applications, hybrid architectures based on LSTM have demonstrated significant effectiveness in sea surface temperature prediction (Hittawe et al., 2022). Similarly, in the traffic management domain, Harrou et al. (2022) demonstrated how LSTM and GRU networks can effectively capture temporal dependencies in traffic flow data, achieving superior forecasting accuracy when combined with appropriate preprocessing techniques. Secondly, the specific features of the presented problem, especially due to the typically limited training data available in industrial settings, are better served by LSTM and GRU architectures than by more complex ones, like Transformers Wei et al. (2023). While recent work has demonstrated good results using Transformer models for time series prediction in different domains like air pollution (Azhahudurai & Veeramanikandan, 2023), wind speed forecasting (Hittawe et al., 2024), and finance (Schäfer et al., 2025), the specific constraints of pharmaceutical equipment, including limited training data availability and requirements for computational efficiency, make LSTM and GRU more suitable choices for our application. These considerations collectively demonstrate that LSTM and GRU models constitute a relevant and effective reference across various domains. Thus, the research on the applicability and on the exploitation of these architectures to the specific challenge of enhancing PP accuracy represents a promising and novel approach with significant potential. By leveraging the ability of these models to capture complex temporal dependencies with a limited amount of data, this study aims to investigate and develop an ADCS capable of improving the precision of PP in pharmaceutical applications.

To achieve our goal, we focus on two specific target volumes: 1.2 ml and 0.3 ml, which represent standard and micro-volume dosing scenarios, respectively, commonly encountered in pharmaceutical filling systems. The 1.2 ml volume was previously investigated in the work based on ARIMA model (Privitera et al., 2023). This volume provides a basis for comparison between the statistical and AI-based approaches. We have selected the 0.3 ml volume to evaluate the performance of these techniques for low-volume dosages, which present unique challenges in terms of accuracy and precision. To enable a comprehensive comparison, we have reproduced the ARIMA approach to include the 0.3 ml target volume, allowing for a thorough evaluation of both methods across different dosage ranges.

All the experiments detailed in this paper have been conducted at Pharma Integration, a company specialized in the development of highly precise and efficient filling lines utilizing robotic technology, in accordance with the recently released Annex 1 of the Good Manufacturing Practices guidelines (European Union, 2022). In summary, this work offers the following major contributions:

- Design and implementation of an ADCS for PP, introducing a detailed hardware-software architecture with two innovative AI-based variants: Online Training and Pre-trained approaches.
- Integration and optimization of LSTM and GRU neural networks as the core predictive engines within our ADCS framework, with systematic parameter tuning to maximize dosing performance particularly for critical volumes below 1.0 ml—a significant challenge in pharmaceutical applications.
- Rigorous comparative evaluation of different ADCS implementations, contrasting our neural network-based approaches with existing ARIMA-based methods through extensive experimental validation on both standard volumes (1.2 ml) and challenging micro-volumes (0.3 ml).

The structure of this paper directly addresses these contributions: Section 2 presents the complete ADCS framework, beginning with a comparative analysis of existing approaches for enhancing industrial system performance (Section 2.1), which critically evaluates different methodologies, their strengths, limitations, and application areas. Following this contextual foundation, we present an overview of the working methodology (Section 2.2), followed by detailed descriptions of the hardware setup (Section 2.3) and software implementation (Section 2.4). We then introduce the core ADCS framework design and operational mechanisms (Section 2.5), and establish the evaluation metrics used to assess system performance (Section 2.6) including our statistical analysis framework that employs confidence intervals and distribution visualizations to validate performance improvements (Section 2.7). Section 3 then develops the implementation and analysis, beginning with a description of the experimental dataset (Section 3.1), followed by our AI-based approaches (Section 3.2). Within this section, we first optimize LSTM and GRU architectures through systematic parameter tuning (Section 3.2.1), then present the two core implementation strategies: the Online Training approach (Section 3.2.2) and the Pre-trained approach (Section 3.2.3). Each implementation includes both a retrospective analysis of model configurations using previously collected datasets without requiring actual machine operation (hereafter referred to as "offline analysis") and a real-world evaluation conducted on the physical system with statistical assessment of performance improvements compared to ARIMA-based methods across different volume ranges (referred to as "online evaluation"). Section 3.3 addresses the practical challenges encountered during deployment, including data quality management, parameter optimization constraints, and computational efficiency considerations that influenced our design decisions. Finally, Section 5 summarizes our findings, discusses their practical implications and limitations for pharmaceutical manufacturing, and suggests promising directions for future research.

2. ADCS setup & Methodology

As previously introduced, this section presents a comprehensive methodology for evaluating and implementing the ADCS. We first examine the landscape of existing approaches for enhancing industrial dosing systems through a critical comparative analysis, identifying their respective strengths, limitations, and application domains. This contextual foundation establishes the methodological progression that led to our current approach and clarifies its potential advantages. Following this comparative framework, we proceed to detail the foundational components required for system deployment, including the industrial dosing system architecture, hardware configuration, and software integration framework. The methodology then introduces the ADCS framework, detailing its operational principles and control mechanisms from a general perspective. Finally, we establish an evaluation framework through precisely defined performance metrics that enable systematic assessment of the system's effectiveness.

2.1. Comparative analysis of approaches for enhancing dosing accuracy

Before going into the specific implementation details of our ADCS, it is valuable to first position this approach within the broader landscape of methodologies aimed at improving precision in manufacturing contexts. Table 1 presents a comparative analysis of the primary approaches identified in literature that reveals a progression of methodologies, each with distinct advantages and limitations when applied to PP accuracy enhancement.

FCC systems represent sophisticated approaches that have demonstrated considerable success across diverse manufacturing domains including end milling, integrated circuit production, and additive manufacturing. These systems excel at compensating for both systematic errors and unpredictable variations through predictive modeling. While traditional FCC implementations rely on detailed mathematical system models, some modern variants adopt a hybrid approach: they use data-driven methods to develop predictive models that, once trained, remain fixed during operation. These approaches still preserve the core FCC philosophy of predictive compensation but avoid the need for explicit physical modeling. However, both traditional and hybrid FCC

Table 1
Comparative analysis of approaches for enhancing industrial system performance.

Approach	Method	Strengths	Limitations	References
FCC	In-process error measurement, stochastic modeling, error prediction and feedforward compensation. Includes both traditional approaches (mathematical modeling) and hybrid variants (data-driven fixed models). **Application-areas:** Milling and grinding applications, IC production, additive manufacturing.	• Compensates for deterministic and stochastic errors • Overcomes time-delay issues through prediction • Enables precision with non-extremely precise machinery • Applicable to various process types	• Requires sensors, precise actuators, and real-time calculations • Precision depends on model accuracy (needs significant tuning) • Traditional approach requires detailed mathematical system representations • Hybrid variants offer limited adaptability to changing conditions • Complex system-specific formulations limit transferability	Francis and Bian (2019), Hu et al. (2020), Wang et al. (2018, 2021), Wu and Ni (1989) and Yan and Huang (2009)
Traditional Mechanical Improvements	Modification of pump physical design: improved rollers, advanced tubing, sophisticated designs. **Application-areas:** Medical and pharmaceutical applications.	• Reduces flow pulsation directly at the source • Improves baseline performance without active control • No additional sensors or complex algorithms required • Direct physical solution to mechanical limitations	• Reduces but does not eliminate pulsation completely • High costs with limited effectiveness • Performance limitations for micro-volumes • Reduced baseline performance when optimizing specific characteristics • Shortened tubing lifespan due to increased mechanical stress	Ferretti et al. (2023), Gasoto et al. (2022), Klespitz and Kovács (2014) and Shieu et al. (2016)
ADCS with ARIMA	Specific implementation using ARIMA model based on feedback from previous dose measurements with black-box approach and continuous runtime parameter adaptation. **Application-areas:** Medical and pharmaceutical applications.	• Software-based compensation, adaptive to pump variations • Improves dosing precision (up to 30%) for 1.2 ml volumes • Continuously adapts model parameters in response to system changes • Applicable to various dosing systems • Fast execution time (<1 s)	• Requires accurate feedback system (precision scales) • Performance depends on auto-regressive window size • Limited parameter capacity for modeling complex relationships • Potential performance limitations with micro-volumes	Privitera et al. (2023)

applications to filling systems present significant challenges. For traditional FCCs, the need to account for the non-linear behaviors of fluid dynamics, tubing elasticity variations, and mechanical wear patterns results in complex models. For hybrid approaches, the static nature of the deployed model limits adaptability to changing operational conditions. The computational and implementation complexity of traditional FCC, coupled with the limited flexibility of hybrid variants, proves prohibitive in pharmaceutical manufacturing environments.

Traditional mechanical improvements, while conceptually straightforward, attempt to address dosing inaccuracies through physical redesign of pump components. This offers the advantage of targeting the error source directly without computational overhead or complex control algorithms. Manufacturers have invested significantly in developing advanced roller mechanisms, specialized tubing materials, and sophisticated pump designs. Nevertheless, these approaches face fundamental limitations. The economic implications are substantial, each incremental improvement in mechanical precision comes with exponentially increasing costs while delivering diminishing returns on accuracy. Furthermore, pharmaceutical applications impose stringent material compatibility requirements that severely constrain design options, often requiring compromise between optimal mechanical properties and biocompatibility or sterilization requirements.

The ADCS with ARIMA models represents a significant methodological advancement by implementing a black-box approach that focuses exclusively on input–output relationships rather than complex internal system dynamics. This innovation fundamentally distinguishes ADCS from traditional FCC systems: instead of requiring precise mathematical modeling of the pump's physical processes, it leverages statistical time-series analysis to predict future behavior based solely on observed dosing patterns. A key strength of this approach is its ability to continuously adapt model parameters at runtime, allowing it to respond dynamically to changing system conditions. The ARIMA implementation has demonstrated that a simple auto-regressive (AR) model can have remarkable efficacy for standard volumes with good computational efficiency, typically executing in less than one second. However, the inherent mathematical constraints of statistical models present potential limitations in capturing complex relationships that may become increasingly significant at micro-volumes, where interactions between fluid properties, mechanical components, and environmental factors create complex temporal patterns. Our current research extends the ADCS framework by investigating the potential of advanced neural network architectures specifically designed for sequential data processing. This approach maintains the fundamental black-box methodology that avoids complex physical modeling while providing expanded parameter capacity and pattern recognition capabilities. These neural architectures can theoretically capture complex non-linear temporal dependencies across varying timescales.

2.2. Black-box approach and experimental platform

Building upon the comparative framework established above, we now describe the specific implementation of our black-box approach which abstracts the dosing system, comprising both the PP and its controller, as a functional unit defined only by its inputs and outputs. This abstraction offers significant adaptability, allowing our methodology to be applied across diverse dispensing devices beyond the specific hardware described in this study.

Industrial filling systems typically integrate filling equipment with weighing scales. Together, these components generate time series data where each point represents a dispensed volume. Analysis of these time series reveals significant correlation between fillings over the process

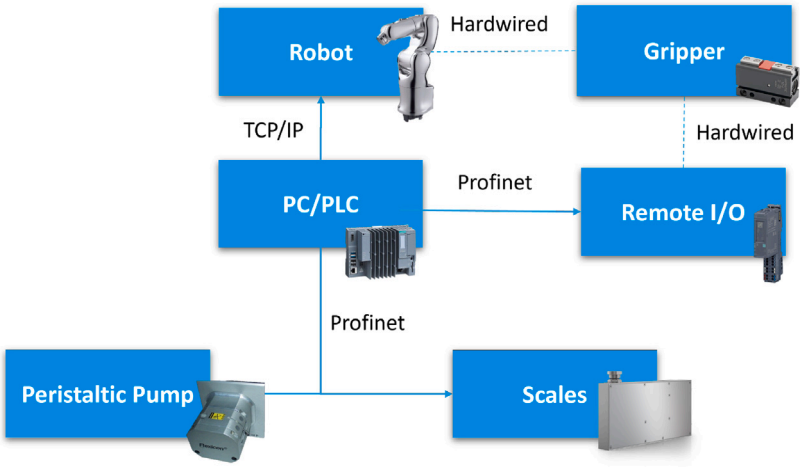

Fig. 2. Experimental workstation architecture and component integration. The system comprises industrial-grade components typically used in pharmaceutical manufacturing. This integrated platform replicates industrial pharmaceutical filling line conditions while enabling controlled experimentation for ADCS validation.

duration (Privitera et al., 2023), creating an opportunity for predictive compensation.

Our implementation interfaces with the pump controller exclusively through well-defined API software interfaces designed to prevent system misuse. This boundary-respecting approach ensures that all data and commands remain within specified operational limits, enhancing process stability while allowing effective compensation.

To validate our methodology across diverse operational conditions, we conducted all experiments using a customized workstation engineered to expedite data collection and simulate various scenarios. This testing platform faithfully emulates the operations of real industrial machinery, incorporating the essential equipment commonly deployed in pharmaceutical manufacturing applications. In the subsequent sections, with reference to Fig. 2, we present a comprehensive description of both the hardware components and software infrastructure that constitute this experimental platform.

2.3. Hardware setup

As described in the previous section, the workstation for the collection of data replicates equipment that is commonly found in an authentic industrial environment. Specifically, we utilized a Flexicon PD12 PP system controlled by a MC100 control unit running firmware version S2 (Jeppesen, 2009) (Watson-Marlow Flexicon, Ringsted, Denmark) for the filling process. To set up the dosing circuit, we linked two sections of pump tubing (Flexicon Accusil™) via a Y-connector before and after the pump. Moreover, we integrated two additional segments of the identical pump tubing including one linked to a liquid reservoir and the other attached to a nozzle, leading to a total piping length of two meters. As reported in Table 2, during the investigation, we analyzed the following configurations:

- For the target volume of 0.3 ml, we employed a nozzle inner diameter of 0.6 mm, tubing inner diameter of 0.5 mm, and Y-connector outer diameter of 2.4 mm.
- For the target volume of 1.2 ml, the setup included a nozzle inner diameter of 1.6 mm, tubing inner diameter of 1.2 mm, and Y-connector outer diameter of 3.6 mm.

In accordance with the PP manufacturer's specifications (Jørgensen, 2008), these setups are intended to handle volumes of [1.0–2.0] ml and [0.01–0.4] ml respectively. For our experiments, volumes of 1.2 ml and 0.3 ml were adopted, which are representative of meaningful real-world dosing quantities. Fig. 3 shows a simplified diagram of the pump tubing setup.

Table 2
PP setup configuration for each target volume: nozzle inner diameter, tubing inner diameter and Y-connector outer diameter.

Target [ml]	Nozzle i.d. [mm]	Tubing i.d. [mm]	Y o.d. [mm]
0.3	0.6	0.5	2.4
1.2	1.6	1.2	3.6

Each test used purified water filtered through a Milli-Q® Advantage A10 system equipped with a Millipak® Express 40 filter. The filling procedures were carried out using the PP at a pre-selected velocity of 600 revolutions per minute (rpm) with an acceleration of 200 rpm/s. The suck-back was set up at 1. Prior to initiating each experiment, a purging session was performed by activating the PP for a few seconds, in order to remove any air from the piping. Following this, we calibrated the PP for a specific fill volume in accordance with the particular experiment. To precisely measure the dispensed volumes, we utilized a Wipotec SL-M 250/300 high-precision scale (Wipotec, Kaiserslautern, Germany). This instrument is characterized by exceptional measurement capabilities, featuring a resolution of 0.002 g, a linearity of ±0.001 g, a repeatability of 0.0005 g, and a settling time of less than 120 ms. Throughout our experimental campaign, in compliance with Good Automated Manufacturing Practice (GAMP) guidelines, we implemented a three-point calibration procedure prior to each experimental session to ensure measurement reliability. The protocol involved verifying the analytical balance at three distinct points using certified reference weights. The calibration sequence began with zero-point verification, followed by measurements using 50 g and 100 g weights. For each calibration point, the displayed value was required to fall within the instrument's specified tolerance limits. This verification process confirmed proper instrument calibration, ensuring measurement accuracy and traceability throughout our experiments. The weighing scale (WS) included an Active Vibration Compensation sensor to reduce interference. Handling of containers was managed by a Denso VS-050S2 robot (Denso, Kariya, Aichi Japan) equipped with a Gimatic MPXM gripper (Gimatic, Brescia, Italy). The system was managed by a Siemens ET 200SP Open Controller (Siemens, Munich, Germany), which acted as a Programmable Logic Controller (PLC) and a PC equipped with an Intel Xeon Gold 6226 and VGPU T4-2Q.

2.4. Software setup

The software structure consists of two main components:

(1) **PLC**: it supervises the functioning of the PP, WS, robot, and gripper.

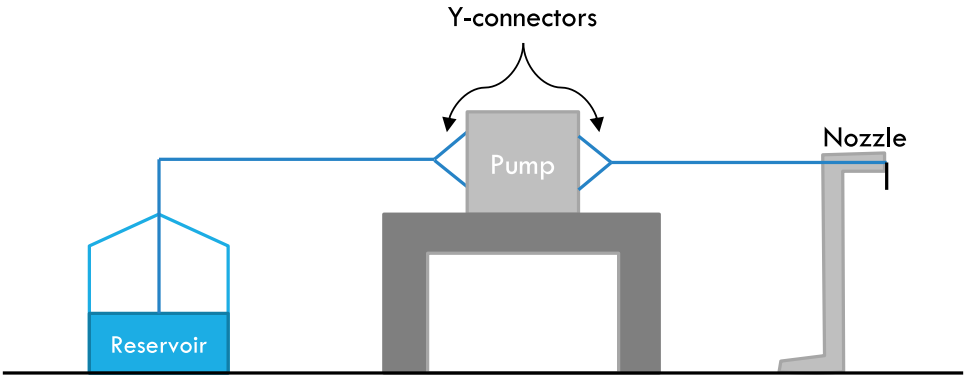

Fig. 3. Peristaltic pump tubing circuit configuration. The system consists of a liquid reservoir, peristaltic pump head, Y-connectors for tubing integration, and dispensing nozzle. The liquid flows from the reservoir through sterile tubing that passes through the pump mechanism, with Y-connectors enabling the connection of multiple tubing segments while maintaining circuit integrity.

(2) **PC**: it executes a Python application that is in charge of the prediction process. It obtains data from the PLC and calculates the next filling volumes.

In further detail, the PLC controls the equipment to replicate a typical filling procedure. The robot is instructed to move a container in order to measure its weight, fill it, and calculate the volume dispensed. To imitate a continuous flow of containers, a loop is implemented in the process, with the procedure described in Algorithm 1.

Algorithm 1 Filling Procedure

```
while Operating do
    while container is not full do
        Place container on the WS
        Store the gross weight
        Place container beneath the needle
        Command the PP to dispense
        Move container back to the WS
        Store the net weight
    end while
    Empty container into the reservoir
end while
```

Depending on the experiment, the PC may run a Python implementation of the ADCS framework detailed in Section 2.5 which receives the dispensed volumes. After accumulating a sufficient quantity, the application implements the compensation model utilizing the `Keras` package (Chollet et al., 2015) with different configurations. The PLC receives the estimated volumes and uses the MC100 to adjust the quantity dispensed.

2.5. ADCS framework

Building upon previous work (Privitera et al., 2023), we introduce the ADCS as a closed-loop compensation framework designed to enhance PP accuracy through predictive adaptation. The framework leverages the hardware and software infrastructure described in the preceding sections to implement a dynamic control strategy and operates through a systematic sequence of operations that integrate measurement, prediction, and compensation phases.

Fig. 4 presents the logical architecture of this framework, which processes four fundamental variables: target volume (V_{tg}), measured dispensed volume (V_d), predicted volume (V_p), and compensation quantity (V_c). The Z^{-1} block represents the one-step delay operator that enables historical data utilization for prediction. This logical architecture materializes in the physical implementation illustrated in Fig. 5, where the control cycle executes as follows:

(A) The PP, handled by the controller MC100, initiates the cycle by dispensing the target volume V_{tg}.

(B) The weighing system measures the actual dispensed volume $V_d(k)$, which the PLC records.

(C) The PLC transmits the dispensed weight to the ADCS. This measurement, combined with the historical sequence $\{V_d(k-1), V_d(k-2), \dots\}$ collected through internal buffering mechanisms, forms the basis for prediction.

(D) The ADCS employs its predictive model to generate $V_p(k)$, forecasting the expected dispensing behavior. Based on this prediction, it computes the compensation quantity $V_c(k)$ required to minimize deviation from V_{tg} and returns is to the PLC.

(E) Finally, the system applies this compensation through the pump controller's `set-volume()` primitive, adjusting the next dispensing operation.

Algorithm 2 ADCS Framework Operation

Require: target volume V_{tg}, training window size TW

```
Initialize empty measurement history H = {}
Calibrate the pump to volume V_tg
while Operating do
    if |H| < TW then
        Dispense target volume V_tg
        Measure dispensed volume V_d(k)
        H ← H ∪ V_d
        Continue
    end if
    Generate prediction V_p(k) using model on H
    Calculate compensation:
```

$$V_c(k) = V_{tg} \cdot \frac{V_{tg}}{V_p(k)} \equiv V_{tg} + \frac{V_{tg}}{V_p(k)} \cdot (V_{tg} - V_p(k)) \tag{1}$$

```
    Apply compensation through set − volume(V_c(k))
    Measure compensated volume V_d(k)
    Update sliding window: H ← (H \ oldest measurement) ∪ V_d(k)
end while
```

This cycle, formalized in Algorithm 2, repeats continuously, with each iteration refining the system's predictive accuracy through accumulated historical data. This algorithmic representation captures the essential components of the framework's operation:

- The initialization phase collecting TW uncompensated measurements ($|H| < TW$).

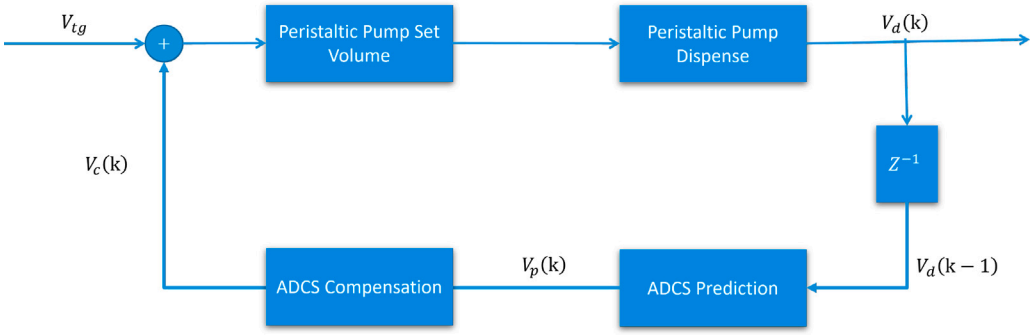

Fig. 4. Logic representation of the implemented ADCS with V_{tg} denoting the target volume, V_d as the dispensed volume, V_p indicating the predicted volume, and V_c representing the compensatory quantity for the next filling. The Z^{-1} block represents a one-step time delay, enabling the system to use historical dispensing data for prediction.

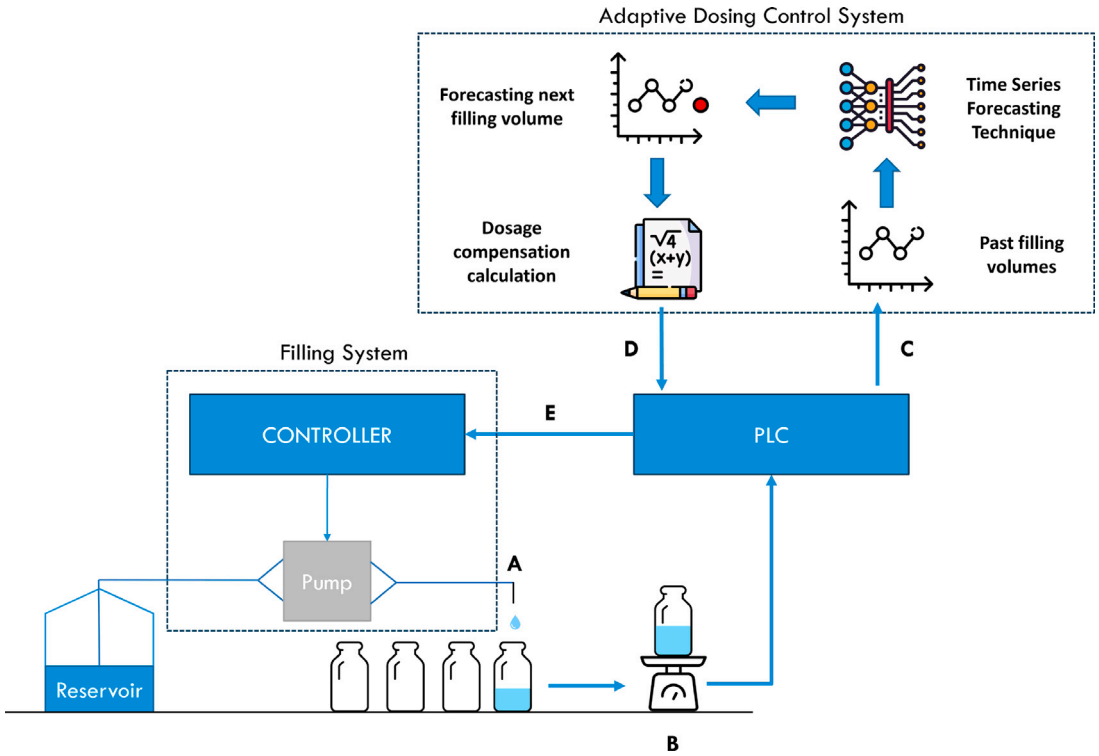

Fig. 5. Dosing control system. (A) The PP is triggered to dispense the volume V_{tg} set into the controller. (B) The filled volume V_d is measured; the weight is registered by the PLC. (C) The PLC transmits the weight to the ADCS which performs the forecasting process. (D) The ADCS returns to the PLC the volume adjustment V_c based on the predicted volume V_p. (E) The PLC sets the adjusted volume into the controller by using the *set-volume*() primitive.

- The prediction step leveraging the chosen model on the current window (V_p).
- The computation of compensation using the ratio-based formula.
- The application of compensation through the controller API.
- The sliding window mechanism maintaining the most recent TW measurements by removing the oldest observation and including the latest compensated volume.

To evaluate the effectiveness of the ADCS framework described in Algorithm 2, we implement a standardized experimental protocol. Each complete experiment, which we name as *subset*, follows the protocol illustrated in Fig. 6 which consists of two distinct operational phases that correspond to specific states of the ADCS framework and utilizes the training window size TW as a key parameter:

(1) **ADCS OFF Phase**: This initial phase corresponds to the initialization stage of Algorithm 2 (lines 1–9). During this phase, the system collects TW uncompensated doses with no active compensation. These measurements serve a dual purpose: they provide the necessary initial data for the predictive model and establish a baseline for system performance evaluation.

(2) **ADCS ON Phase**: This subsequent phase implements the main operational loop of Algorithm 2 (lines 10–14). In our experimental design, this phase extends for $2 \cdot TW$ doses, during which the ADCS actively applies compensations based on predictions generated by the selected model. The *Operating* condition expressed in line 3 remains true until we reach the predetermined limit of $2 \cdot TW$ compensated doses, at which point the experimental evaluation concludes.

The extended duration of the ADCS ON phase to twice the duration of the ADCS OFF phase allows us to observe the system's behavior even after the initial data window has been fully populated with compensated measurements, providing insight into the long-term stability and effectiveness of the compensation strategy. To ensure robust evaluation, we collected multiple subsets for each volume range and

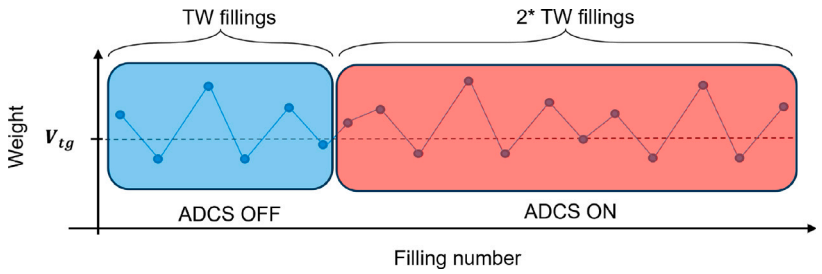

Fig. 6. Experimental protocol phases for ADCS evaluation. The process comprises two distinct phases: an initial phase (ADCS OFF) of length TW that provides baseline measurements, followed by a compensation phase (ADCS ON) of length $2 \cdot TW$ that enables evaluation of the system's performance with active compensation. The dashed line represents the target volume V_{tg}, while blue and red points indicate uncompensated and compensated measurements, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

compensation approach. This experimental protocol remains consistent across all ADCS implementations discussed in this paper, enabling direct comparison between different predictive approaches and allowing for statistical analysis across subsets.

Central to the framework's effectiveness is the volume compensation formula shown in Eq. (1). This implements an inverse-ratio compensation strategy that adjusts the dispensed volume based on the predicted deviation from the target. The formula can be interpreted in two equivalent forms: as a ratio-based scaling of the target volume or as an additive correction term. In its multiplicative form ($V_c = V_{tg} \cdot \frac{V_{tg}}{V_p}$), the equation scales the target volume by the ratio of desired to predicted volumes. When the prediction indicates potential under-dispensing ($V_p < V_{tg}$), this ratio becomes greater than unity, increasing the commanded volume. Conversely, predicted over-dispensing ($V_p > V_{tg}$) results in a ratio less than unity, reducing the commanded volume. The additive representation ($V_c = V_{tg} + \frac{V_{tg}}{V_p} \cdot (V_{tg} - V_p)$) explicitly shows the correction term as proportional to the predicted error ($V_{tg} - V_p$), scaled by the ratio $\frac{V_{tg}}{V_p}$. This formulation highlights how the compensation magnitude adapts to both the size of the predicted error and the relative scale of the target and predicted volumes.

The modular nature of this framework allows for the integration of various predictive models while maintaining the fundamental compensation mechanism. Different prediction strategies can be implemented by modifying the model used in Step 10, while the overall control loop structure remains substantially unchanged. In Section 3.2, we discuss specific implementations of this framework leveraging artificial neural networks.

2.6. Evaluation metrics

The quantitative assessment of ADCS performance necessitates a comprehensive evaluation framework that captures both the precision and accuracy aspects of the dosing system. To this end, we employed two key metrics: Standard Deviation (STD) and Root Mean Square Error (RMSE).

STD serves as a measure of precision, quantifying the dispersion of fill volumes around their mean. In our context, a lower value indicates higher consistency in dosing, which is critical for maintaining uniform product quality and ensuring patient safety. The STD is mathematically expressed as:

$$STD = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(V_d(k) - \mu)^2},$$

where $V_d(k)$ represents the kth observed fill volume, μ is the mean fill volume, and n is the total number of observations.

RMSE, on the other hand, is a measure of accuracy, reflecting how close the predicted values are to the actual observed values. This serves as our primary metric for evaluating model performance on previously collected datasets, measuring how accurately a predictive model can reconstruct dosing patterns from historical data. This offline assessment, performed on extensive datasets gathered during experimental campaigns, enables us to evaluate different model configurations without requiring actual machine operation. This offline assessment is crucial for identifying promising model configurations before real-world implementation. It is expressed mathematically as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(V_p(k) - V_d(k))^2}, \tag{2}$$

where $V_p(k)$ represents the kth predicted fill volume. RMSE was selected instead of other metrics, like Mean Absolute Error, to place greater emphasis on values that considerably differed from the predictions. This choice is important in pharmaceutical applications, where large deviations from the target dose can have significant clinical implications.

For real-world system evaluation, we adopt also a slightly different RMSE formulation which summarizes the error between the actual dispensed volumes $V_d(k)$ and the target volume V_{tg}. Consequently, for the online tests conducted on the real system, we transition from the standard RMSE to a modified version, called $RMSE_{PP}$, that serves as our accuracy indicator and which is calculated as follows:

$$RMSE_{PP} = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(V_d(k) - V_{tg})^2} \tag{3}$$

To illustrate the practical significance of $RMSE_{PP}$ in pharmaceutical applications, Fig. 7 presents two contrasting dosing scenarios. Scenario A exhibits high precision (low dispersion) but poor accuracy, with measurements consistently deviating from the target volume V_{tg}. Conversely, Scenario B demonstrates lower precision (higher dispersion) but superior accuracy, with measurements distributed more widely but properly centered around the target volume.

This comparison reveals an important consideration in pharmaceutical dosing: traditional precision metrics like STD would favor Scenario A due to its lower dispersion, despite its systematic deviation from the target volume. In contrast, $RMSE_{PP}$ correctly identifies Scenario B as superior for pharmaceutical applications, where delivering the prescribed target volume is the primary objective, even if accompanied by slightly higher variability.

This version of RMSE, captures both systematic errors and random variations in the dosing process, offering a detail view of the compensated system's performance. This approach aligns more closely with the practical objectives of our ADCS, focusing on the actual dispensed volumes in relation to the target, rather than the intermediate predictions, which are conversely useful for gaining insight into the reasons and mechanisms behind the overall results. The combination of these metrics provides a robust evaluation framework: RMSE enables efficient offline screening of model configurations, STD assesses the consistency of dosing performance, and $RMSE_{PP}$ measures real-world system accuracy. Together, they ensure a thorough understanding of system behavior from initial model selection through to operational deployment.

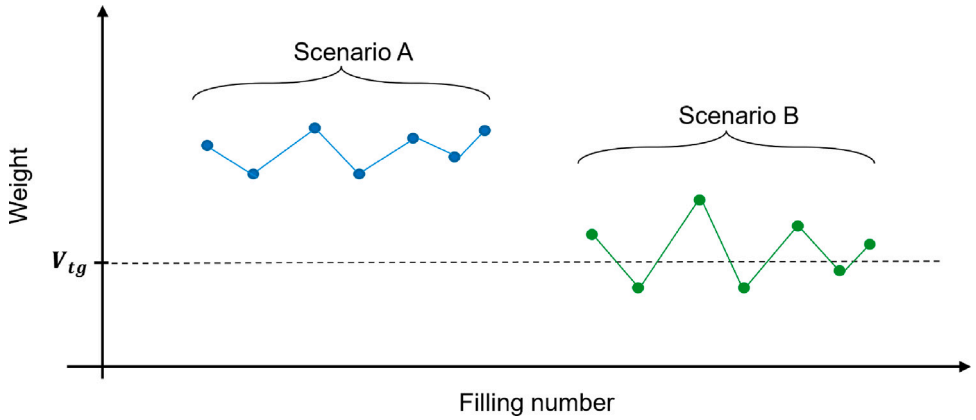

Fig. 7. Comparison of two filling scenarios illustrating the distinction between precision and accuracy in dosing systems. Scenario A demonstrates high precision but low accuracy, with measurements tightly clustered but consistently offset from the target volume (V_{tg}). Scenario B shows comparatively lower precision but superior accuracy, with measurements appropriately centered around the target despite greater dispersion. This visualization highlights why $RMSE_{PP}$ provides a more relevant assessment for pharmaceutical applications than metrics focused solely on dispersion.

Based on these evaluation tools, we now need a method to quantify the effectiveness of our compensation approach. By comparing them between the ADCS OFF and ADCS ON phases, we can quantitatively evaluate the improvement achieved through the compensation mechanism. For clarity, we define the performance improvements in terms of gain for each metric, calculated as:

$$\text{Gain}_{metric} = \left(1 - \frac{\text{metric}_{ADCS_{ON}}}{\text{metric}_{ADCS_{OFF}}}\right) \times 100\% \tag{4}$$

where $\text{metric}_{ADCS_{OFF}}$ represents the value of STD or RMSE_{PP} when ADCS is inactive (uncompensated dosing), and $\text{metric}_{ADCS_{ON}}$ represents the corresponding value when ADCS is active (compensated dosing). A positive Gain value indicates improvement (reduction in error or variation), while a negative value indicates performance deterioration. For each ADCS implementation approach and volume configuration, we collected multiple independent subsets to ensure robustness. The performance gains reported in our results tables represent the means of these multiple runs.

2.7. Statistical analysis framework

To ensure the significance and reliability of the performance improvements reported in this study, we implemented a statistical validation framework that quantifies the confidence in our results and provides a rigorous foundation for comparative analysis between different compensation methods.

Our statistical analysis framework consists of two complementary components:

(1) **Confidence Intervals**: We calculated 95% confidence intervals for all key performance metrics to quantify the uncertainty associated with our results. These intervals provide a range within which we can expect the true performance improvement to lie with 95% confidence. Using the Gain metric defined above, for each configuration we collected several subsets of $3 \cdot TW$ doses each. If we denote the number of subsets as n_s, the mean gain across these runs as $\bar{x}$, and the sample STD of these gain values as s, the confidence interval is calculated as:

$$\bar{x} \pm t_{\alpha/2,n_s-1} \cdot \frac{s}{\sqrt{n_s}} \tag{5}$$

where $t_{\alpha/2,n_s-1}$ is the critical value from Student's t-distribution with $n_s - 1$ degrees of freedom at significance level $\alpha = 0.05$. Importantly, confidence intervals that do not include zero indicate statistically significant changes (equivalent to $p < 0.05$ in hypothesis testing), while intervals containing zero represent non-significant effects.

(2) **Distribution Visualization**: We employed violin plots to visualize the complete distribution of errors (difference between dispensed volume and target volume) before and after compensation. These visualizations reveal not only changes in central tendency but also modifications in the shape, spread, and potential multimodality of the error distributions, providing a more comprehensive understanding of system behavior than summary statistics alone.

This statistical approach was systematically applied to our experimental data, enabling rigorous validation of performance improvements and comparative analysis of different compensation techniques. By incorporating these methods, we ensure that the performance enhancements attributed to our ADCS implementations are statistically significant and not the result of random variations or experimental artifacts.

3. Experiments & Results

Having established the foundational components of our ADCS framework, we now proceed to demonstrate its practical implementation and effectiveness through a systematic experimental campaign. The following sections describe how this methodological framework was applied to real-world dosing scenarios, revealing both the capabilities and limitations of our approach across different volume ranges and implementation strategies. We present a comprehensive investigation structured in three complementary parts. First, we examine the detailed operational characteristics of the PP without compensation, establishing baseline performance and identifying underlying behavioral patterns. Second, we evaluate two distinct AI-based implementation strategies for prediction and compensation — Online Training and Pre-trained approaches — comparing their effectiveness in improving dosing accuracy across different volume ranges. Finally, we summarize the practical challenges encountered during implementation, discussing data quality management, parameter optimization constraints, and computational efficiency considerations that influenced our design decisions.

3.1. Experimental dataset and pump dosing behavior

The initial experiments aimed to understand the operational dynamics of the PP by running it continuously for several days, collecting a significant amount of data. According to the target volume, each batch started by assembling the components of the PP (tubing, Y-connector, nozzle) described in Table 2, followed by a calibration of the pump to the designated target volume. The calibration was performed using

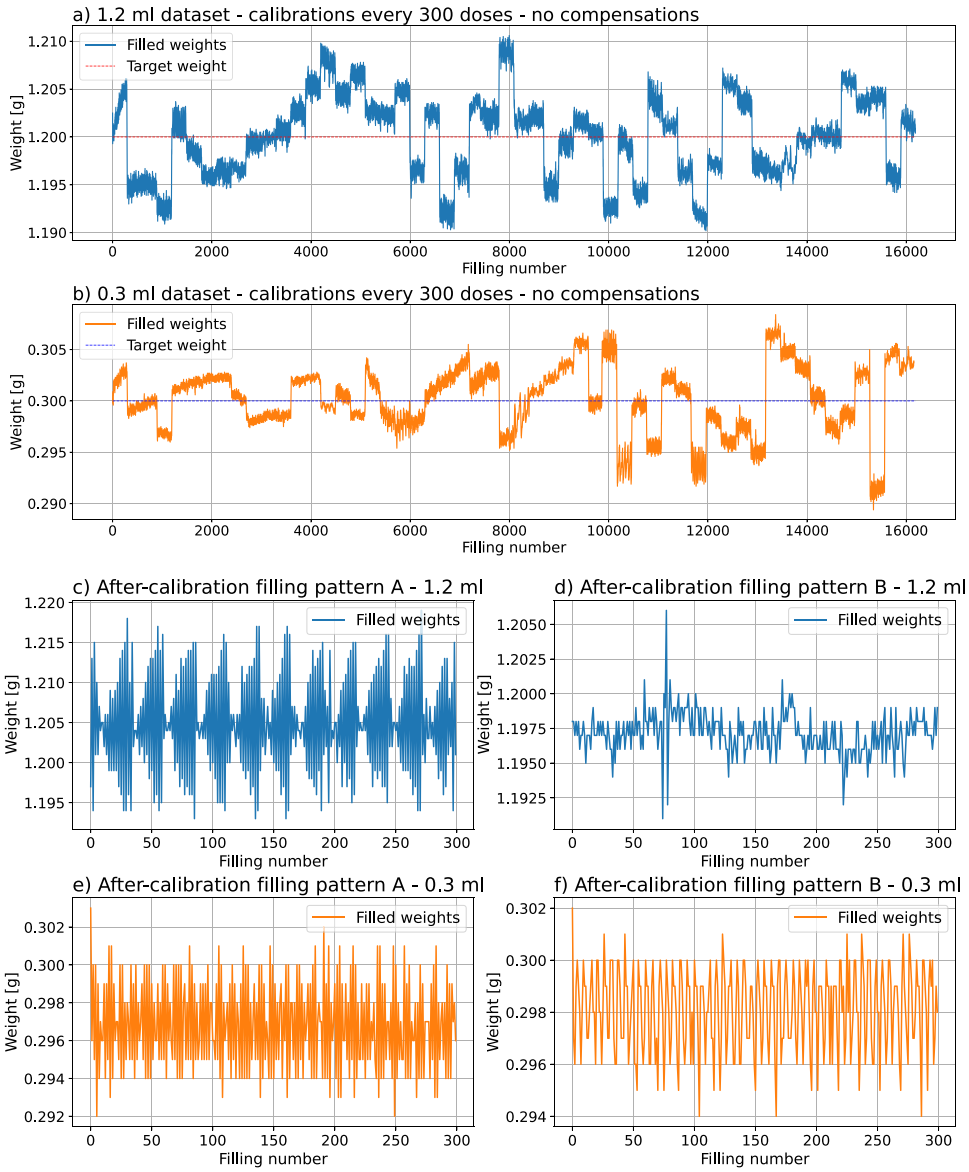

Fig. 8. Dataset for (a) 1.2 ml and (b) 0.3 ml obtained through calibration of the PP every 300 doses. The different offsets caused by the calibration process are clearly visible. Divergent time patterns are observed following calibration in subsets (c) and (d) which correspond to a volume of 1.2 ml. Similar differences in time series patterns are observed in subsets (e) and (f), which correspond to a volume of 0.3 ml.

the dedicated command provided by the pump controller's API. This straightforward command allows calibrating the pump by specifying the actual dispensed volume compared to the target one, enabling the pump to automatically adjust its internal parameters to achieve more precise dosing. The initial procedure included a purging phase to eliminate air from the tubes, followed by five calibration iterations to ensure accurate dispensing. As discussed, for this study we focused on two target volumes, 1.2 ml and 0.3 ml, collecting approximately 16,000 doses for each volume. The batches consisted of several subsets of 300 doses each. The pump was re-calibrated at the beginning of each subset using the same API command mentioned earlier, but performing only a single calibration iteration to investigate how the operational parameters of the PP evolved after calibration. Temperature regulation, crucial for precision (Privitera et al., 2023), was not strictly monitored due to re-calibration every 300 doses, which adjusts for product temperature. This experimental design, with periodic calibration and substantial data collection, provides a comprehensive view of the pump's behavior under real operating conditions.

As shown in Fig. 8(a) and (b), each calibration phase produces an output characterized by distinct offsets across the collected data. Furthermore, on closer inspection of certain random subsets, it can be confirmed that not only do the offsets change, but the temporal patterns produced by the device also show variation (see Fig. 8(c), (d), (e), and (f)).

3.2. ADCS implementations

Building upon the ADCS framework presented in Section 2.5, we now present two specific implementations using AI models to enhance pump accuracy. These implementations exploit the procedure described in Algorithm 2 to realize the framework's predictive compensation capabilities. Specifically, we explore the use of LSTM and GRU networks for predicting upcoming fill volumes.

In time series analysis, AI models provide great capacity to capture complex relationships within the data. These models are trained on sequential data, enabling them to retain and utilize information from past observations. Unlike traditional recurrent neural networks, GRU and LSTM architectures incorporate gating mechanisms that control the flow of information, allowing them to selectively remember or forget past states.

The LSTM model is characterized by its ability to capture long-term dependencies in the data. It achieves this by maintaining a cell state, which can store information over extended sequences, allowing it to capture trends or patterns that span significant time intervals. Additionally, LSTM networks possess gate units that manage the flow of information, including the input gate (which controls new information), the forget gate (which manages the retention of information from the previous state), and the output gate (which regulates the information passed to the output).

Similarly, GRU models also incorporate gating units but have a more streamlined architecture compared to LSTM. GRU merge the memory and hidden state into a single hidden state vector, simplifying the computational process. This streamlined design can be advantageous in scenarios with limited computational resources.

Both models excel at modeling sequences with intricate dependencies, making them well-suited for time series forecasting tasks. They are particularly effective in scenarios where the data exhibits long-term trends, seasonal patterns, or complex temporal relationships. Additionally, these models can adapt to changing patterns in the data, allowing them to provide accurate predictions even in dynamic environments.

In our study, we developed and evaluated two distinct implementation strategies of the ADCS framework:

1. Online Training Approach: discussed in Section 3.2.2, it continuously updates the model based on recent measurements, allowing for real-time adaptation to changing system conditions.
2. Pre-trained Approach: analyzed in Section 3.2.3, it utilizes models trained on historical data, offering faster execution times at the cost of reduced adaptability.

In the subsequent sections, we present an evaluation methodology that integrates offline analysis and real-world validation to assess our proposed strategies. This employs a two-phase approach wherein we first conduct an offline evaluation utilizing the dataset collected to determine optimal model configurations for practical deployment. We approached this by analyzing the prediction accuracy of neural architectures across different structural configurations, training them on sequential dosing data to predict the subsequent volume based on a window of preceding measurements. This analysis includes the evaluation of configurations of the number of network layers (Section 3.2.1), followed by hyper-parameter optimization including neuron count and temporal window size in the context of both Online Training (Section 3.2.2) and Pre-trained approaches (Section 3.2.3). Through this methodology we establish empirical performance benchmarks that support the subsequent real-world implementation phase, where we validate the optimized configurations in actual dosing scenarios collecting between 16 and 31 independent experimental subsets per configuration as detailed in Table 4. This methodical repetition across multiple operational conditions enables meaningful confidence interval calculations and provides strong statistical evidence for the performance improvements observed.

3.2.1. LSTM and GRU general architecture optimization and parameter tuning

The initial phase of this study focused on determining the optimal network configuration for the presented models. This foundational investigation set the stage for subsequent experiments detailed in the following sections. Unless otherwise specified, the procedure delineated in this paragraph corresponds to the configuration used during the experiments.

The analysis included both the 1.2 ml and 0.3 ml dataset described in Section 3.1. A critical pre-processing step involved using the `StandardScaler` from the `Sklearn` library (Pedregosa et al., 2011) to standardize individual doses and normalize them. This step ensured consistency in the data representation and enabled model performance comparison between different volumes.

To facilitate the training of the neural network, we introduced the time-window parameter p which dictated the number of preceding dose measurements included in each sample. The selection of this parameter was guided by both theoretical considerations and empirical observations from previous work where using the AR(10) model — that is, an auto-regressive model with a working window of 10 —(Privitera et al., 2023), it was established that 10 previous measurements provided sufficient historical context for effective prediction in standard volumes. Given that neural architectures like LSTM and GRU can theoretically capture more complex temporal dependencies than linear auto-regressive models, we hypothesized that these networks might benefit from longer windows of historical data. To systematically evaluate this hypothesis, we expanded our investigation to include three distinct time-window values: $p = [10, 20, 30]$. Consequently, each sample was formed as a vector of p continuous dose measurements, with the subsequent dose value serving as the target for prediction. For instance, assuming D_i represents the ith dose within the current subset composed by n elements, the sample-target pairs took the following form:

$$\begin{aligned} &(D_1, D_2, D_3, \dots, D_p), && D_{p+1} \\ &(D_2, D_3, D_4, \dots, D_{p+1}), && D_{p+2} \\ &\dots \\ &(D_{n-p}, D_{n-p+1}, D_{n-p+2}, \dots, D_{n-1}), && D_n. \end{aligned} \tag{6}$$

Through this process, we generated a comprehensive set of sample-target pairs. Taking into account $n = 300$ and $p = [10, 20, 30]$, the training set was formed by assigning 85% of the samples. This split ratio, which differs slightly from traditional proportions, was chosen after initial experiments to better address the specific challenges of our industrial application. While our dataset represents a significant collection for the pharmaceutical manufacturing sector, it differs in scale from the massive datasets available in other domains. The 85-15 split provides a better balance between having sufficient data to capture complex temporal patterns during training while maintaining a meaningful test set. Within the training set, we reserved 15% of the data as a validation set. The algorithm utilized a *tanh* activation function and employed batch learning with updates after every eight samples (batch size). For model training, we employed the Adam optimizer. Additionally, the maximum allowable number of epochs during training was limited to 600. After a 100-epoch warm-up period, we implemented an early stopping criterion with a patience parameter of 20 epochs. Training was terminated if the cost function calculated on the validation set showed minimal variation over this consecutive period.

In order to determine the ideal number of layers required for precise forecasts within the models, a thorough investigation was carried out. The objective of this phase was to establish the layer arrangement that can appropriately accommodate the characteristics of the dataset. Together with the previously outlined methodology, we conducted several tests, systematically varying the number of layers while keeping a constant number of neurons per layer. It is worth noting that the choice of the number of neurons was arbitrary, serving as a placeholder to assess if increasing the number of layers could yield any discernible impact on predictive performance. This study investigated a range of neural network designs from a single layer to a maximum of five layers. Given that altering the number of neurons yielded no discernible effects, the results of these experiments, where the number of neurons per layer is eight, are illustrated in Fig. 9 highlighting the efficacy of a single-layer configuration.

While this investigation has set various foundational parameters, further optimization remains necessary for each specific model implementation. The subsequent sections present detailed analyses where we systematically evaluate model-specific parameters, including the optimal number of neurons within the single layer, the impact of different time-window lengths, and in certain cases, the effect of various input feature configurations. These targeted investigations aim to fine-tune each model's architecture to maximize its predictive capabilities while maintaining computational efficiency.

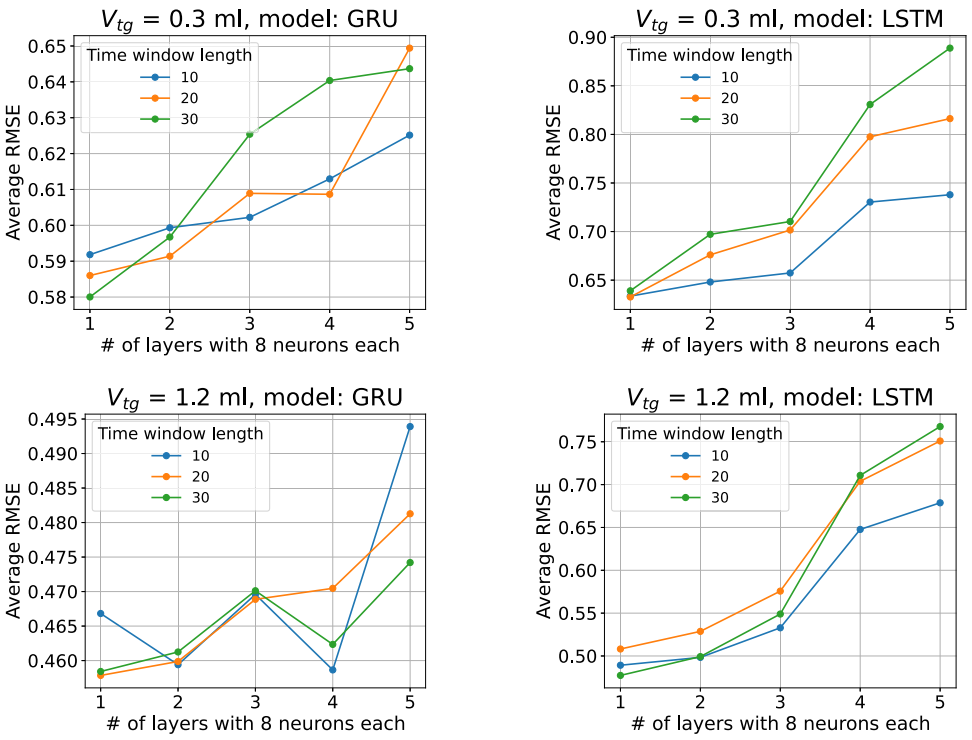

Fig. 9. Analysis of LSTM and GRU model performance for 0.3 ml and 1.2 ml with regard to the number of layers. The model's success rate is determined by the average value across multiple subsets of RMSE calculated from the test set. A consistent trend was observed across all tested models and time-window parameters, with a single layer emerging as the optimal choice.

3.2.2. ADCS implementation: Online training approach

Building upon the ADCS framework presented in Section 2.5, this implementation addresses the dynamic nature of PP behavior through continuous model adaptation. The approach leverages the observation that the pump's statistical properties evolve not only after calibration, as discussed in Section 3.1, but also following each compensation through the controller's primitive *set-volume*() (Privitera et al., 2023) which allows the system to dispense a specific volume, effectively modifying the dosing process from the established calibration point. This dynamic behavior necessitates an iterative training procedure that continuously updates the predictive model.

While maintaining the core ADCS structure, the Online Training approach implements the prediction phase through a sliding window mechanism, where the model is retrained at each step using the most recent observations. Algorithm 3 (see Fig. 10) formalizes this implementation, highlighting the key differences from the general framework.

Algorithm 3 ADCS with Online Training

Require: training window TW, history window size p

1: Initialize empty measurement history $H = \{\}$
2: Initialize compensated doses counter $n = 0$
3: Calibrate the pump to the target V_{tg}
4: Collect TW fillings in H with no compensations
5: **while** $n < 2 \cdot TW$ **do**
6: Train the AI model on top of H (Fig. 10a)
7: Generate prediction using model on last p elements of H (Fig. 10b)
8: Calculate compensation according to Eq. (1)
9: Apply compensation
10: Measure compensated volume (Fig. 10c)
11: Update sliding window (Fig. 10d)
12: $n \leftarrow n + 1$
13: **end while**

The uncompensated doses collected in the initialization phase (Step 4) serve a dual purpose, as described in Section 2.5: they provide the initial training data and establish a baseline for system performance evaluation following the standard experimental protocol. Applying this protocol, we evaluate the system's performance by comparing metrics between the ADCS OFF and ADCS ON phases.

Offline analysis. Prior to implementing this process in a real test, it is crucial at this point to identify the optimal configurations for the employed AI models in terms of number of neurons and time-window length. This necessitated another training session on the datasets collected and presented in Fig. 8. The training process adhered to the pre-processing methodology outlined in Section 3.2.1, where sample-target pairs were generated for each subset of $n = 300$ doses. Notably, each subset was treated as an independent dataset. The models employed included both GRU and LSTM single layer architectures with a number of neurons ranging from 1 to 30, applied to 0.3 ml and 1.2 ml volumes. The results of this training, expressed in terms of the standard metric $RMSE$ reported in Eq. (2), are depicted in Fig. 11. Model performance is quantified as the mean RMSE across various subsets of the test set, yielding the following key observations:

1. Both models yield remarkably similar results. Given that LSTM entails a higher number of parameters compared to GRU, the latter is evidently the preferred choice.
2. The network's performance exhibits swift improvement when transitioning from 1 to approximately 6 neurons. However, the enhancement becomes notably slower or non-existent when moving from 6 to 30 neurons.
3. The network's performance demonstrates minimal variation with respect to the time-window length p. Across values of 10, 20, and 30, no substantial differences are discernible. Hence, an optimal choice is a time-window of 20.

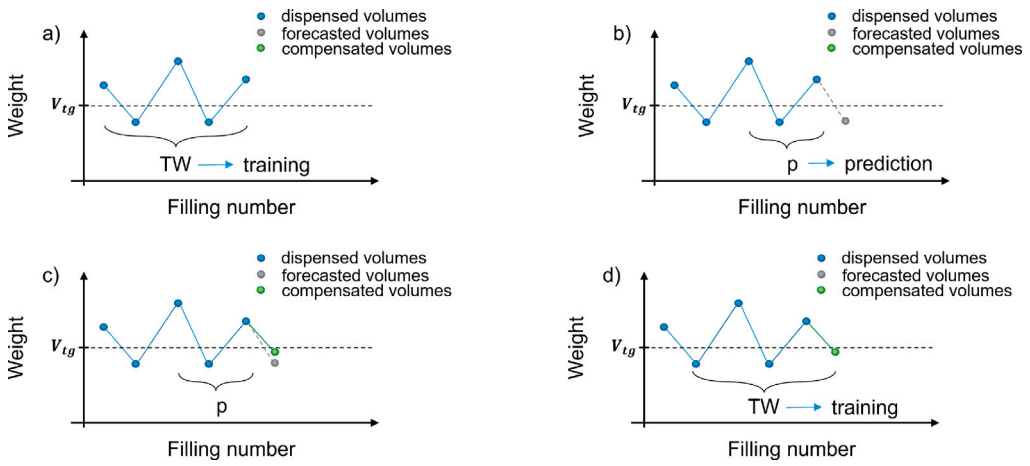

Fig. 10. ADCS with Online Training representation. (a) TW dispensed volumes are collected and the AI model trained. (b) Considering the last p doses, the forthcoming fill is forecasted. (c) PP is compensated according to the forecasted volume and the next fill is triggered resulting in a dispensed volume closer to the V_{tg}. (d) The TW is moved to consider the last compensated volume and the AI model is trained again.

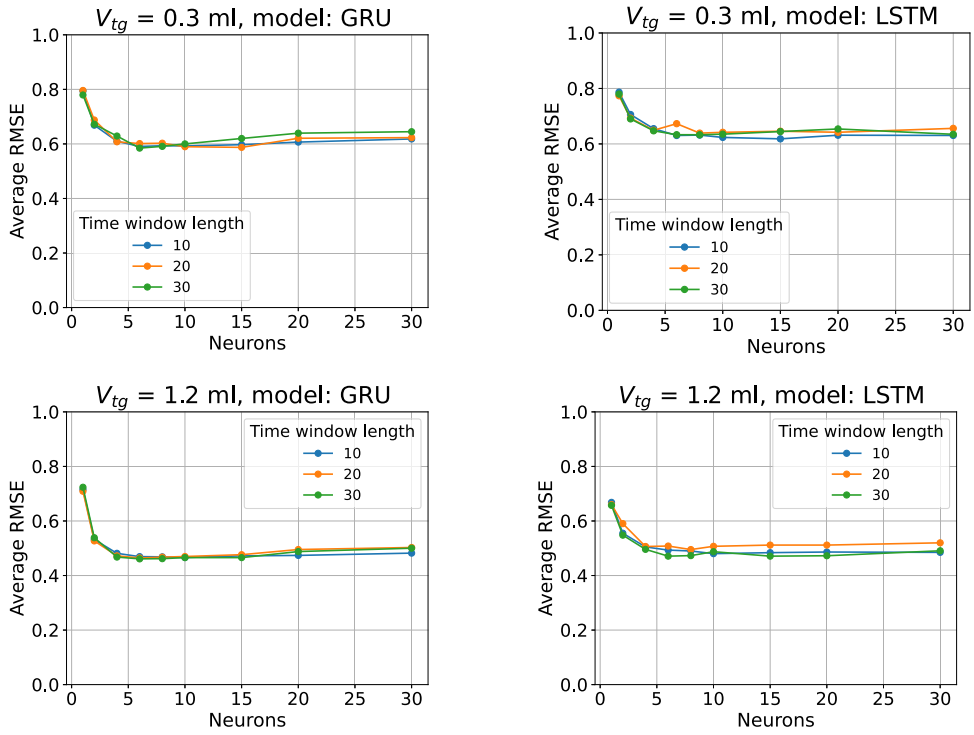

Fig. 11. Analysis of LSTM and GRU model performance for 0.3 ml and 1.2 ml with regard to the number of neurons in the intermediate layer and the length of the time-window p. The model's success rate is determined by the average value across multiple subsets of RMSE calculated from the test set.

Online analysis. Drawing from insights obtained through these analysis, we conducted an extensive experimental validation of the strategy formalized in Algorithm 3. The experimental campaign consisted of several independent test iterations using the workstation architecture detailed in Fig. 2. This methodical repetition enabled robust statistical analysis of the compensation system's performance across multiple operational conditions. For all experimental runs, we employed a GRU model with a minimal yet effective architecture: single layer, 6 neurons, and a time-window parameter $p = 20$.

Different TW lengths were explored to train the model's parameters, representing the quantity of data points used for training, extracted from available time series data. The TW selection significantly impacts model accuracy and its ability to detect data patterns. A larger TW captures long-term dependencies, while a smaller one focuses on short-term variations. The ideal TW size varies based on the task; here, a range of 100 to 300 points was assessed to balance capturing data dynamics and maintaining a concise TW.

Results reported in Table 4 highlight success with a TW of 100 and V_{tg} of 0.3 ml. Comparing the uncompensated and compensated data segments, significant improvements were observed, with gains of 31.6% in STD and 38.4% in $RMSE_{PP}$. Statistical analysis confirms the significance of these improvements, with 95% confidence intervals of [25.0%–38.3%] for STD gain and [32.7%–44.2%] for RMSE_{PP} gain. The fact that these intervals do not include zero provides strong statistical evidence that our approach delivers consistent improvements across different operational conditions. The relatively narrow confidence intervals, particularly for RMSE_{PP}, indicate high reliability in the performance enhancement achieved. As illustrated in Fig. 12, the distribution visualization further confirms these findings. The violin plot for Online Training with $TW = 100$ at 0.3 ml demonstrates a

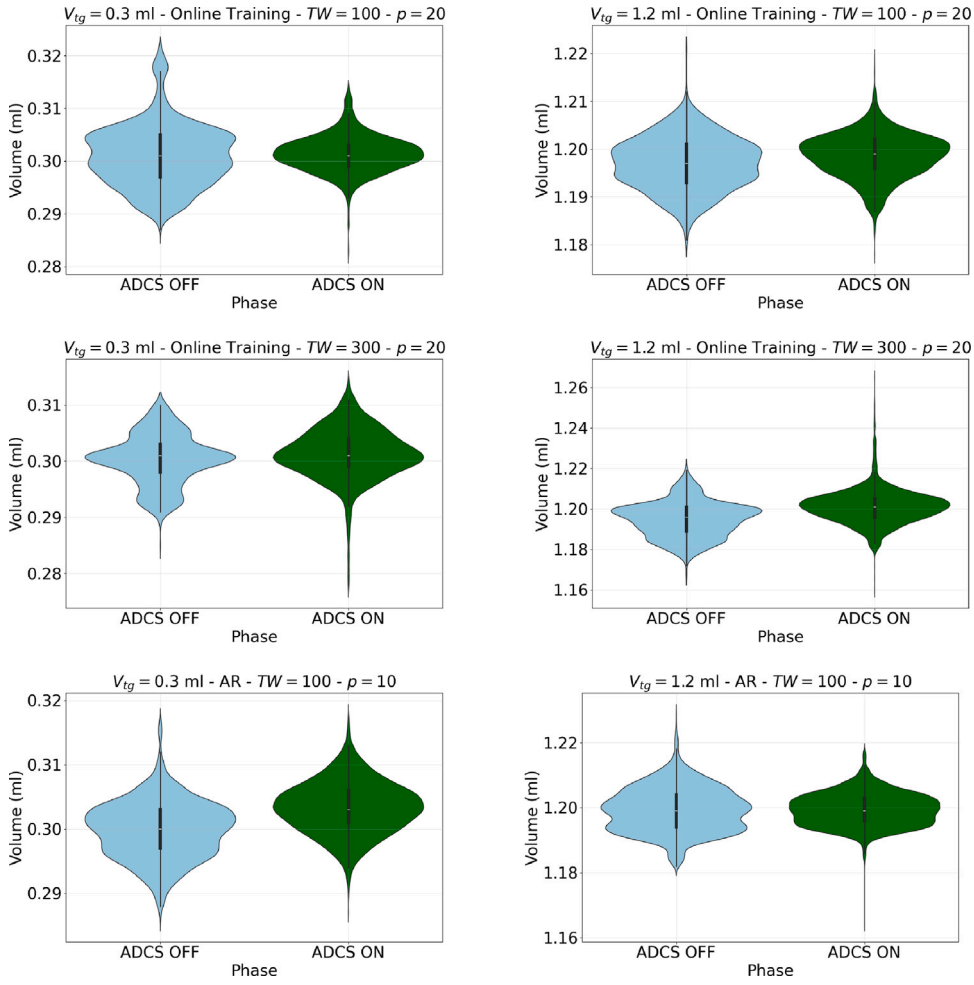

Fig. 12. Error distribution visualization comparing ADCS OFF and ADCS ON phases for AR and Online Training approaches.

remarkable transformation from a multi-modal distribution with pronounced variability to a more compact, unimodal distribution tightly centered around the target volume. This structural reshaping of the entire distribution profile visually confirms both the improved accuracy and precision quantified by our statistical metrics.

For V_{tg} of 1.2 ml instead a negative impact on STD has been observed, with an average loss of 5.7% but the 95% confidence interval for this metric [−21.9%–10.4%] includes zero, indicating that the observed decrease in precision is not statistically significant. This suggests that the precision neither consistently improves nor deteriorates across different operational conditions, but rather varies depending on specific circumstances. On the other hand, applying the ADCS results in a positive average gain of 23.2% with respect to the $RMSE_{PP}$, with a confidence interval of [14.3%–32.1%] that remains consistently positive. This statistically significant improvement in RMSE_{PP} demonstrates that despite the variable effect on precision, the system consistently achieves its primary objective of centering the dispensed volumes closer to the target value. The violin plot for 1.2 ml with TW = 100 in Fig. 12 visually confirms this analysis, showing a distribution that maintains similar width (reflecting the non-significant STD change) but with a notable center shift closer to the target volume. While the improvement for 1.2 ml appears more subtle than for 0.3 ml when visualized, the statistical significance of the RMSE_{PP} gain demonstrates that even this less dramatic distributional change represents a meaningful accuracy enhancement for larger volumes.

After conducting experiments with TW of 300, it became evident that the results were not favorable for both volumes. Statistical analysis of these configurations revealed concerning performance, particularly for the 1.2 ml volume where we observed a statistically significant negative impact on STD with a confidence interval of [−46.9% to −13.9%]. For the 0.3 ml volume, both metrics showed non-significant results with confidence intervals including zero, further confirming the lack of consistent improvement with this configuration. The violin plot visualizations in Fig. 12 provide clear evidence supporting these statistical findings. For the 1.2 ml volume, the ADCS ON distribution exhibits substantially increased dispersion compared to ADCS OFF, with a notable expansion toward values higher than the target, representing a marked deterioration in precision. For the 0.3 ml volume, the ADCS ON distribution achieves a structural transformation from multimodal to unimodale, suggesting some positive reshaping of the underlying process dynamics. However, this structural improvement is counterbalanced by more extended tails in both directions and fails to achieve the significant overall narrowing observed with $TW = 100$. This mixed effect explains the statistically non-significant results for this configuration, indicating that while the model with $TW = 300$ does restructure the distribution in potentially beneficial ways, these benefits are offset by increased occurrences of extreme values. This implies that training the network with excessively large TW values may lead to inaccuracies during the prediction process, likely due to the overwhelming amount of historical data being considered. This underlines the importance of carefully selecting an appropriate TW size to strike a balance between capturing relevant patterns and avoiding information overload. Additionally, it is noteworthy that the execution time of the model, which is calculated by averaging the time required across multiple subsets to perform the Step 6 of Algorithm 3, exhibits sensitivity to the chosen TW. For instance, with a TW of 100, the

model completes its prediction step in about 10 s. In contrast, when employing a TW of 300, the average execution time surpasses 20 s.

In the field of industrial procedures, computational efficiency is crucial. Within the context where our experiments were conducted, a computation time of 10 s is compatible with the machine cycle. However, it should be considered that cycle times may vary in different industrial settings. In such cases, hardware advancements are available to optimize performance. Recent improvements in GPU technology have led to more powerful accelerators, resulting in more efficient training times for AI models (Peccerillo et al., 2022).

To provide a comprehensive comparison, we also consider the results from previous work on ADCS using the ARIMA model (Privitera et al., 2023). This study demonstrates that the AR component of the ARIMA model is particularly effective when applied to PP. Specifically, an AR(10) model which follows a training methodology identical to that described in Algorithm 3, showed promising results for the 1.2 ml target volume with a $TW = 100$. To ensure completeness, we re-tested the model for the 1.2 ml volume and we extended this approach to the 0.3 ml target volume, maintaining the same TW. Each dataset comprised over 5000 clean entries. The PP setup followed the components described in Table 2, maintaining consistent pump parameters across both configurations, including 600 rpm speed and 200 rpm/s acceleration. For data collection, the pump was re-calibrated every 300 fillings, with each cycle consisting of 100 fillings without ADCS followed by 200 fillings with ADCS active. The results for both volumes are included in Table 4.

The AR(10) model demonstrates strong performance, particularly in terms of execution time, which is below 1 s for both volumes we tested. While the previous work does not explicitly report execution times, focusing instead on $RMSE_{PP}$ and STD metrics, we can make a reasonable inference about the computational efficiency. Given that the AR(10) model structure remains consistent regardless of the target, and that the model complexity is primarily determined by the fixed window size of 10 and the TW value, it is logical to expect that the execution time for other volumes would also be under 1 s.

Regarding performance metrics, as reported in the previous work (Privitera et al., 2023) for the 1.2 ml target, the AR(10) model achieved a 19.5% improvement in STD and a 29.4% improvement in $RMSE_{PP}$. In our replication of the 1.2 ml test, we observed slightly different results, with an STD of 13.4% and an $RMSE_{PP}$ of 27.2%. Analysis confirms that both improvements are statistically significant, with confidence intervals of [3.3%–16.8%] for STD and [20.5%–33.8%] for RMSE_{PP}. As visualized in Fig. 12, the distribution comparison demonstrates this improvement clearly, with the ADCS ON phase showing a notably narrower distribution centered closer to the target volume of 1.2 ml compared to the wider distribution observed in the ADCS OFF phase. For the 0.3 ml target, our results showed a −8.7% change in STD but a positive 7.3% improvement in $RMSE_{PP}$. Interestingly, while the STD decrease is not statistically significant with a confidence interval of [−19.4%–2.0%], the RMSE_{PP} improvement does reach statistical significance with a confidence interval of [0.2%–14.4%], albeit with a lower bound very close to zero. This narrow confidence interval approaching zero suggests a marginally significant effect with considerable variability across experimental runs. The distribution visualization in Fig. 12 provides additional insight into this phenomenon: the ADCS ON distribution shows a reduction in extreme values, particularly in the lower tail, while simultaneously shifting the distribution's central mass slightly above the target. This pattern indicates that the AR(10) model at micro-volumes achieves its modest RMSE_{PP} improvement primarily by eliminating extreme deviations rather than by improving overall centering. Such behavior is characteristic of the challenges encountered at micro-volumes, where even slight variations represent a substantial percentage of the target volume, making consistent dosing control particularly difficult. This limitation of the AR(10) model for micro-volumes further supports our finding that AI-based approaches like GRU with Online Training are substantially more effective for volumes below 1.0 ml.

The slight differences observed between the original study and our replication for the 1.2 ml target volume are worth noting. In particular, the $RMSE_{PP}$ values are quite consistent. The more noticeable difference is in the STD values that could be attributed to several factors, with the initial conditions of the system being a primary consideration. The PP's performance can be influenced by factors such as number of subsets collected, tube wear, product temperature or differences in the setup of the pump circuit, all of which may have differed between the two experimental setups. Despite these differences, both sets of results confirm the effectiveness and robustness of the AR(10) model in improving dosing accuracy.

The contrasting results for the 0.3 ml target volume, where we observed a slight decrease in STD but an improvement in $RMSE_{PP}$, highlight the challenges in maintaining consistent performance across different volume ranges. This suggests that while the AR(10) model is generally effective, its performance may vary depending on the target volume and specific system conditions.

These results highlight the efficiency and effectiveness of statistical models for larger volumes, providing a valuable benchmark for comparison with the AI-based approaches. They also emphasize the importance of replicating experiments across different conditions to fully understand the robustness and adaptability of dosing compensation models.

3.2.3. ADCS implementation: Pre-trained approach

While the Online Training method described before demonstrates significant effectiveness, it is important to assess the difference between its compensation capabilities and a pre-trained mode where training is done beforehand and conceptually once, and the trained system does inference at runtime. This approach can be potentially promising when extremely fast inference times could be needed in other industrial contexts, and gives the opportunity to assess how much of the improvement of the online-one derives from the adaptation of the continuous training and how much from the global input predictability in itself. Notably, while implemented within our ADCS framework, this Pre-trained approach conceptually bridges toward the hybrid FCC methodologies discussed in Section 2.1. Specifically, building upon the ADCS framework presented in Section 2.5, this implementation replaces the iterative training mechanism with a pre-trained model approach. This shift offers the key advantage of developing a general pump model applicable across various calibration conditions, eliminating the need for repetitive training. It leverages the vast parameter capacity of neural networks to effectively capture the system's complexity. Algorithm 4 (see Fig. 13) formalizes this implementation with the key difference in the prediction mechanism: instead of retraining at each step, the model maintains a fixed set of weights, obtained through the offline training session described later, while adapting to new measurements through the sliding window mechanism. The feature vector X represents the collection of input features that are populated according to the different methodologies discussed in the next paragraph. Considering the same protocol used in the Online Training approach, each test iteration follows the ADCS OFF/ON phases shown in Fig. 6, enabling direct comparison.

Offline analysis. The initial phase involved an offline training session aimed at determining optimal model configurations. The training was conducted similarly to the subset tests discussed in Section 3.2.1. However, we determined the sample-target pairs described in Eq. (6) by analyzing the entire dataset as a whole, instead of treating each subset as a distinct dataset. Additionally, we implemented a bigger batch size of 32, an extended maximum epoch limit of 1200, and a deferred early-stop criterion after 300 epochs. These modifications were made to improve the overall convergence of the model parameters across the entire dataset. When researching the most effective input configuration for the network, we studied three distinct methodologies:

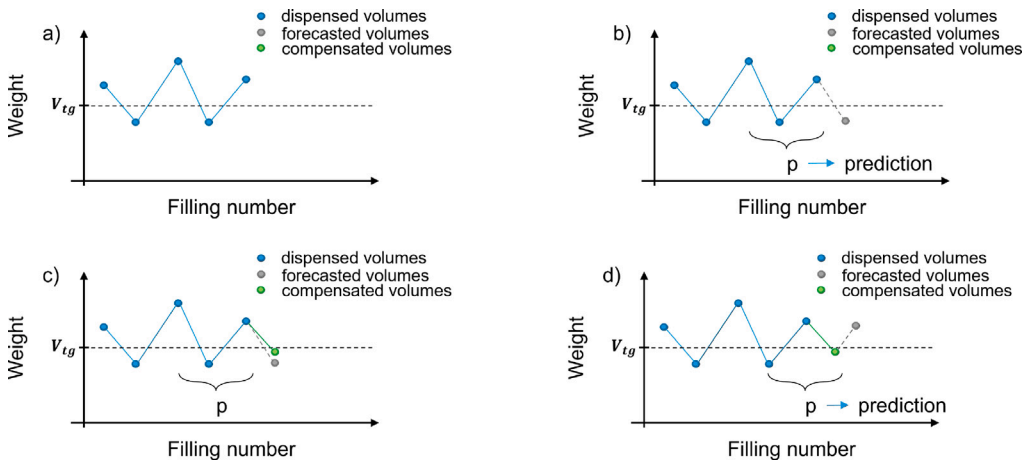

Fig. 13. ADCS with Pre-Training representation. (a) Some dispensed volumes are collected. (b) Model processes feature representation of last p measurements to forecast next volume. (c) PP is compensated based on prediction, applying compensation to achieve convergence toward target volume V_{tg}. (d) The p window is moved to consider the last features and the AI model is asked again for a prediction.

Algorithm 4 ADCS with Pre-trained Approach

Require: window TW, history window size p, model type m

Initialize feature vectors $X = \{\}$
Initialize compensated doses counter $n = 0$
Calibrate the pump to the target V_{tg}
Collect TW fillings with no compensations and update X (Fig. 13a)
while $n < 2 \cdot TW$ **do**
    Generate prediction using m on last p elements of X (Fig. 13b)
    Calculate compensation according to Eq. (1)
    Apply compensation
    Measure compensated volume (Fig. 13c)
    Update sliding window and feature vectors X (Fig. 13d)
    $n \leftarrow n + 1$
end while

1. **one-feature:** This method utilized a singular feature derived from the time series representing dosed volumes, emphasizing the straightforward consideration of temporal dosing information. Given a time-window p, the training set is composed of pairs $(\mathbf{X}_i, y_i)$ where:

$$\mathbf{X}_i = \begin{bmatrix} D_i \\ D_{i+1} \\ \vdots \\ D_{i+p-1} \end{bmatrix} \in \mathbb{R}^p$$

$$y_i = D_{i+p}$$

for $i = 1, \ldots, N - p + 1$, where N is the total number of doses in the dataset and D_i represents the ith dosed volume.
For instance, given a dataset of doses $[1.2, 1.3, 1.1, 1.4, 1.2]$ and a time-window $p = 3$, the two input-target pairs would be structured as:

$$\mathbf{X}_1 = \begin{bmatrix} 1.2 & 1.3 & 1.1 \end{bmatrix}, \quad y_1 = 1.4$$
$$\mathbf{X}_2 = \begin{bmatrix} 1.3 & 1.1 & 1.4 \end{bmatrix}, \quad y_2 = 1.2$$

2. **two-features:** A subsequent extension involved adding a second feature after transforming the original dosed volume series into a sequence of pairs $(\mathbf{X}_i, y_i)$. This new feature, represented by the rolling mean, captured the progressive trend within each input window while resetting for each new group. Given a time-window p, the training set is composed of pairs $(\mathbf{X}_i, y_i)$ where:

$$\mathbf{X}_i = \begin{bmatrix} D_i & \mu_{1,i} \\ D_{i+1} & \mu_{2,i} \\ \vdots & \vdots \\ D_{i+p-1} & \mu_{p,i} \end{bmatrix} \in \mathbb{R}^{p\times 2}$$

$$y_i = D_{i+p}$$

for $i = 1, \ldots, N - p + 1$, where N is the total number of doses in the dataset. The rolling mean $\mu_{t,i}$ for each time step $t = 1, \ldots, p$ in the window $\mathbf{X}_i$ is calculated as:

$$\mu_{t,i} = \frac{1}{t} \sum_{k=1}^{t} D_{i+k-1},$$

where D_{i+k-1} represents the kth element of the subsequence $\mathbf{X}_i$. This ensures that the rolling mean resets for each new input window. Continuing with the previous example sequence and $p = 3$, the two input-target pairs would incorporate the rolling means as follows:

$$\mathbf{X}_1 = \begin{bmatrix} 1.2 & 1.20 \\ 1.3 & 1.25 \\ 1.1 & 1.20 \end{bmatrix}, \quad y_1 = 1.4$$
$$\mathbf{X}_2 = \begin{bmatrix} 1.3 & 1.30 \\ 1.1 & 1.20 \\ 1.4 & 1.27 \end{bmatrix}, \quad y_2 = 1.2$$

where, for instance, in $\mathbf{X}_1$ the means are calculated as $\mu_{1,1} = 1.2$, $\mu_{2,1} = (1.2 + 1.3)/2 = 1.25$, and $\mu_{3,1} = (1.2 + 1.3 + 1.1)/3 = 1.20$. Note how the mean calculation resets for $\mathbf{X}_2$.

3. **three-features:** Building upon the two-features approach, this method introduced a third feature in addition to the dosed volumes and their rolling mean represented by the compensated volume V_c dispatched to the pump with each dosing. This enhanced approach aimed to capture the nuanced dynamics of the dosing system by incorporating compensatory information as a relevant predictor in the model. The training set is composed of pairs $(\mathbf{X}_i, y_i)$ where:

$$\mathbf{X}_i = \begin{bmatrix} D_i & \mu_{1,i} & V_{c,i} \\ D_{i+1} & \mu_{2,i} & V_{c,i+1} \\ \vdots & \vdots & \vdots \\ D_{i+p-1} & \mu_{p,i} & V_{c,i+p-1} \end{bmatrix} \in \mathbb{R}^{p\times 3}$$

$$y_i = D_{i+p}$$

for $i = 1, \ldots, N-p+1$, where N is the total number of doses in the dataset, $\mu_{t,i}$ represents the rolling mean as defined previously, and $V_{c,i}$ represents the compensation volume at time step i.

Table 3
Optimal model configurations based on the offline analysis for Pre-traine approach.

Target [ml]	Method	Model	Neurons	p
1.2	Two-features	GRU	8	30
1.2	Three-features	GRU	10	20
0.3	Two-features	LSTM	8	30
0.3	Three-features	GRU	10	30

Extending the previous example with compensation volumes $V_c = [0.1, 0.2, 0.15, 0.1]$, the first input-target pair would be:

$$\mathbf{X}_1 = \begin{bmatrix} 1.2 & 1.20 & 0.1 \\ 1.3 & 1.25 & 0.2 \\ 1.1 & 1.20 & 0.15 \end{bmatrix}, \quad y_1 = 1.4$$

maintaining the same rolling mean computation while incorporating the compensation information.

It is important to note that the three-features method required distinct datasets. To train the network, it was necessary to use data where volumes were actively compensated. This was achieved using the datasets based on the AR(10) model mentioned in the previous section, collected for both the 0.3 ml and 1.2 ml target volumes. The data collection process, as previously described, involved cycles of 300 fillings, consisting of 100 fillings without ADCS followed by 200 fillings with ADCS active. It is noteworthy that for the final dataset used in the three-features method, only the ADCS-active fillings were included. This means that the subsets of 100 uncompensated fillings from each subset were removed, ensuring that the dataset exclusively contained actively compensated data. Subsequently, we conducted a training session for all the methods, with the three-features employing the same methodology as the other techniques, but utilizing this specially prepared dataset. The obtained results are presented in the following figures, divided by volume: Fig. 14 displays results for the 0.3 ml volume, while Fig. 15 presents results for the 1.2 ml volume.

Each subplot presents the average RMSE of the network for a specific combination of factors: model type, volume, and number of features. The number of features used in the model is represented by the rows: one-feature (top row), two-features (middle row), and three-features (bottom row). The x-axis represents the number of neurons in the network, while the y-axis represents the average RMSE.

The predictive performance of the networks shows a slight improvement when transitioning from the one-feature method to the two-features method. For the 0.3 ml dataset, a notable observation is the nearly constant RMSE value with varying numbers of neurons in the network (left subplot). Based on the results obtained, we selected the two-features method as a valid approach for further testing in the online phase for both volumes. The three-features method deserves separate consideration. Despite a higher RMSE, this was also chosen for further online analysis. The motivation behind this choice is twofold: first, the dataset used to train the model is not the same as that used in the previous methods, making direct comparisons among the performance of these approaches potentially misleading. Second, there is a promising prospect of a model that inherently incorporates the third feature representing the compensated volume, adding valuable depth to the overall model.

Upon examining the plots, we identified the optimal configurations of the models based on the offline analysis for Pre-trained approach. The configurations, summarized in Table 3, include the type of recurrent neural network (GRU or LSTM), the number of neurons in the intermediate layer, and the length of the time-window parameter p for both 1.2 ml and 0.3 ml volumes.

Online analysis. As previously described, two techniques were chosen for each volume, namely, the two-features based on the rolling mean approach and the three-features approach. The succeeding online testing phase targets the assessment of these models in an actual setting. During the evaluation phase, we assessed system performance through multiple metrics: STD for precision measurement, $RMSE_{PP}$ for accuracy assessment, and execution time for computational efficiency. Each configuration was tested applying Algorithm 4 and the ADCS OFF/ON protocol shown in Fig. 6, enabling direct comparison across multiple experimental runs with results averaged to provide robust performance indicators.

The findings, displayed in Table 4, demonstrate a negative result for the two-features approach, both for 0.3 ml and 1.2 ml, with statistically significant deterioration in precision. The confidence intervals for STD gain ([−33.9% to −8.3%] for 0.3 ml and [−22.7% to −1.1%] for 1.2 ml) clearly exclude zero, confirming that this precision loss is consistent across experimental conditions. Regarding accuracy, while the observed effects on $RMSE_{PP}$ were negative for 0.3 ml (−2.8%) and positive for 1.2 ml (3.3%), neither reached statistical significance as indicated by confidence intervals ([−9.5%–3.9%] and [−2.7%–9.3%] respectively). The distribution visualizations in Fig. 16 provide nuanced insights into these statistical findings. For the 0.3 ml volume, the ADCS ON distribution exhibits a subtle but noticeable widening compared to ADCS OFF, along with a slight shift of its peak toward values above the target, aligning with both the negative STD gain and negative $RMSE_{PP}$ effect. For the 1.2 ml volume, the transformation is more complex: the ADCS ON distribution develops a more intricate structure with an apparent indentation around 1.19 ml suggesting the emergence of bimodality, and shows increased asymmetry with a more pronounced lower tail. This altered distribution shape visually confirms the statistically significant precision loss observed in our metrics, while the maintenance of a similar central position explains why the positive $RMSE_{PP}$ effect (3.3%) fails to reach statistical significance despite its numerical improvement.

The three-features approach exhibits promising characteristics, particularly for improving accuracy. For the 0.3 ml scenario, although the approach resulted in a negative STD gain (−6.3%) with a confidence interval of [−16.2%–3.6%], it achieved a statistically significant improvement in RMSE_{PP} (9.2%) with a confidence interval of [1.6%–16.8%]. The performance at larger volumes proved even more encouraging, with the 1.2 ml configuration yielding statistically significant positive gains in both metrics: 4.6% improvement in STD with a confidence interval of [0.1%–9.1%] and 13.8% in RMSE_{PP} with a confidence interval of [8.4%–19.3%]. Fig. 16 provides visual confirmation of these statistical findings with revealing distributional details. For the 0.3 ml volume, the ADCS ON distribution demonstrates a transformation from a slightly asymmetric shape to a more symmetrical profile that is better centered on the target volume. While maintaining similar overall width (explaining the non-significant STD change), the distribution shows notably improved centering and uniformity, especially in the upper region around 0.30–0.31 ml, confirming the significant RMSE_{PP} improvement. For the 1.2 ml volume, the transformation is even more pronounced, with the ADCS ON distribution showing both enhanced centering and a visible narrowing, particularly evident in the reduction of the upper tail (above 1.21 ml). This dual improvement in both distribution position and compactness provides a clear visual parallel to the statistically significant gains observed in both accuracy and precision metrics, highlighting the effectiveness of incorporating compensation volume as a third feature, especially for larger dosage ranges.

As expected, there has been a notable improvement in operational efficiency, with an average prediction time of less than 1 s for both volumes, demonstrating the computational advantage of the Pre-trained approach over the Online Training method.

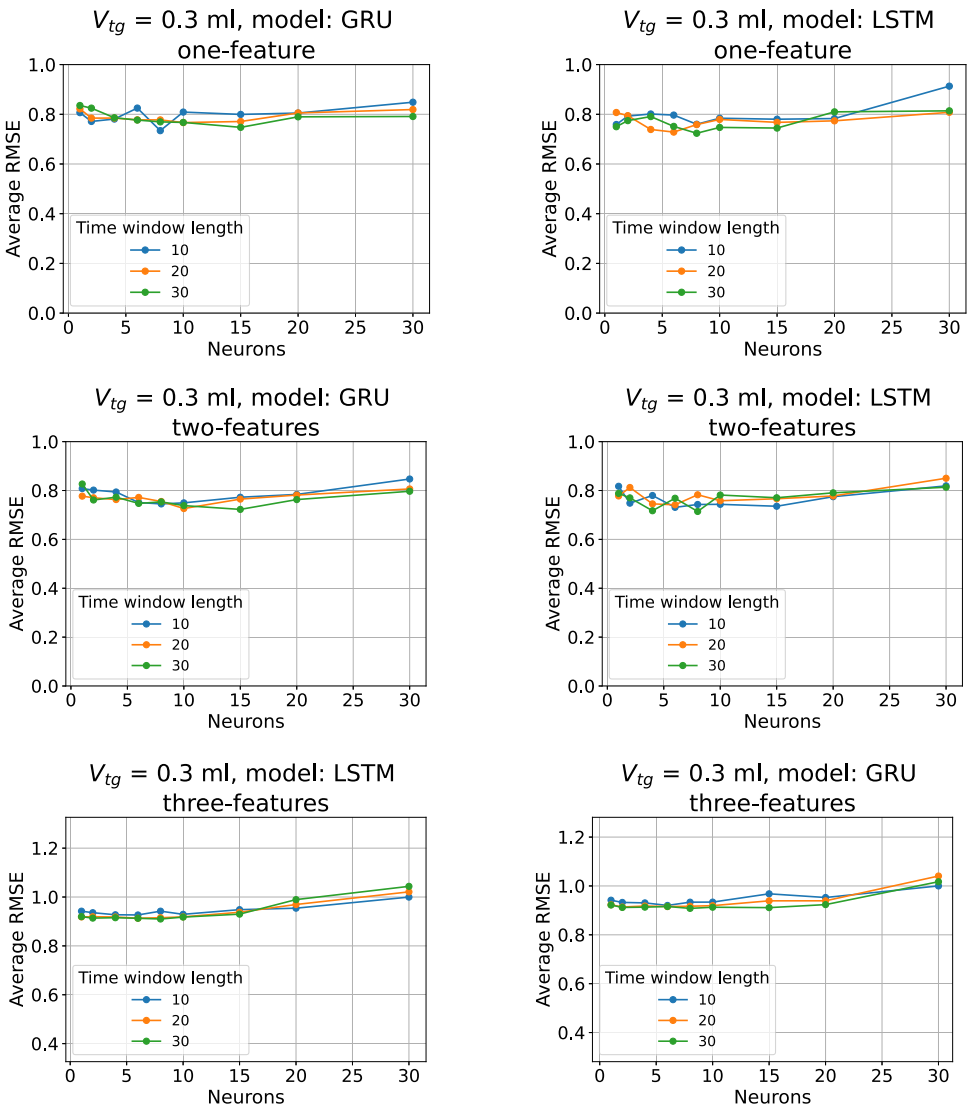

Fig. 14. Performance analysis of LSTM and GRU models for the 0.3 ml dataset, examining the influence of the number of neurons in the intermediate layer and the length of the time-window p for the different proposed methods. Model success is evaluated based on the average RMSE across multiple subsets calculated from the test set.

Table 4
Precision gain of the different ADCS approaches using STD, $RMSE_{PP}$ and execution time as the evaluation metrics. Positive gain indicates increase in performance. Values include 95% confidence intervals [lower - upper]. The Subsets column indicates the number of independent experimental runs each composed by $3 \cdot TW$ doses. For small volumes online training performs better, while for bigger volumes AR appears the most balanced choice, with online-training as a valid alternative if pursuing accuracy.

Target [ml]	Method	TW	p	Subset	STD Gain [%]	RMSE ∗ PP Gain [%]	Time [s]
0.3	AR	100	10	31	−8.7 [−19.4–2.0]	7.3 [0.2–14.4]*	<1
0.3	Online-training	100	20	19	31.6 [25.0–38.3]*	38.4 [32.7–44.2]*	11
0.3	Online-training	300	20	17	−9.4 [−20.5–1.7]	4.7 [−7.3–16.7]	25
0.3	Two-features	–	30	26	−21.1 [−33.9 to −8.3]*	−2.8 [−9.5–3.9]	<1
0.3	Three-features	–	30	30	−6.3 [−16.2–3.6]	9.2 [1.6–16.8]*	<1
1.2	AR	100	10	28	13.4 [3.3–16.8]*	27.2 [20.5–33.8]*	<1
1.2	AR (previous study)	100	10	–	19.5	29.4	–
1.2	Online-training	100	20	16	−5.7 [−21.9–10.4]	23.2 [14.3–32.1]*	12
1.2	Online-training	300	20	17	−30.4 [−46.9 to −13.9]*	−5.9 [−18.6–6.8]	27
1.2	Two-features	–	30	24	−11.9 [−22.7 to −1.1]*	3.3 [−2.7–9.3]	<1
1.2	Three-features	–	20	30	4.6 [0.1–9.1]*	13.8 [8.4–19.3]*	<1

* Statistically significant improvements (CI does not include zero).

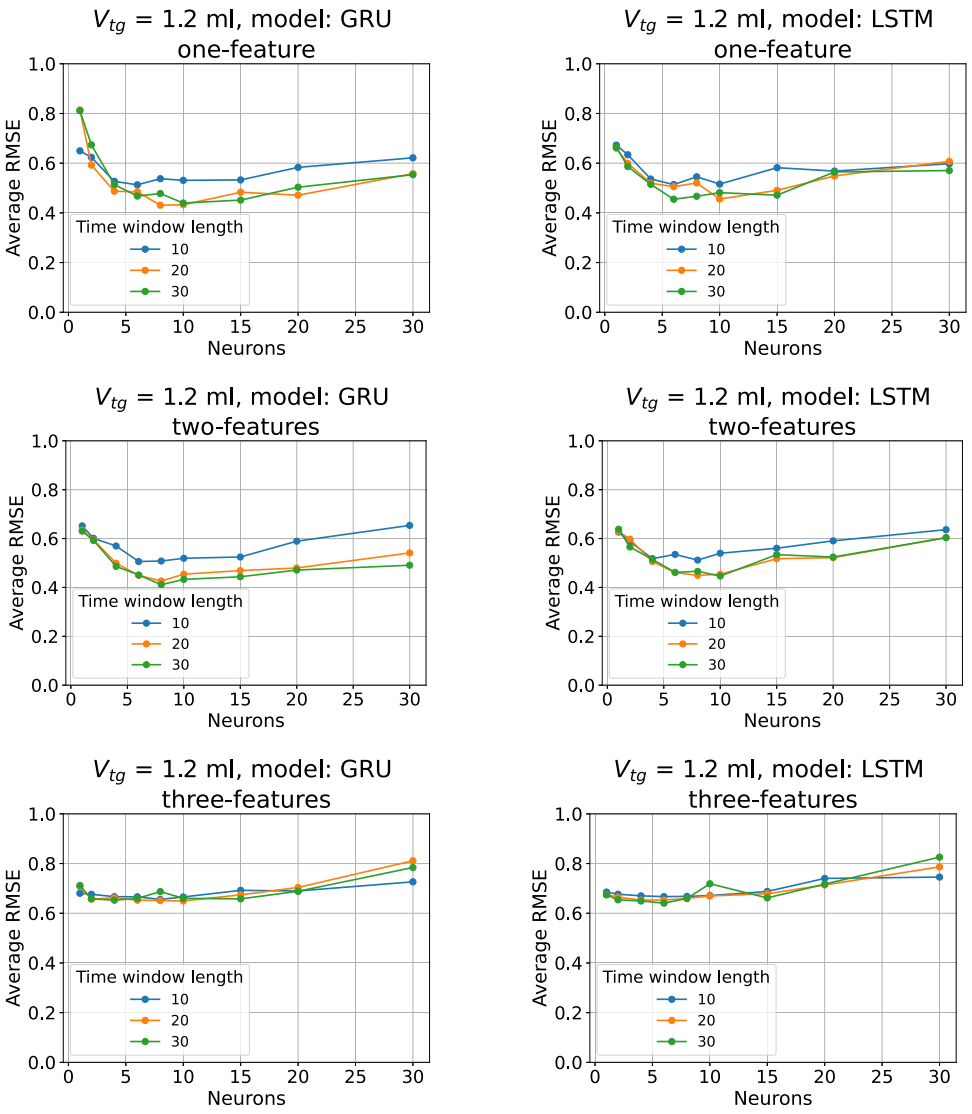

Fig. 15. Performance analysis of LSTM and GRU models for the 1.2 ml dataset, examining the influence of the number of neurons in the intermediate layer and the length of the time-window p for the different proposed methods. Model success is evaluated based on the average RMSE across multiple subsets calculated from the test set.

3.3. Implementation challenges and practical considerations

The experimental results presented thus far demonstrate the potential of both Online Training and Pre-trained approaches for enhancing peristaltic pump dosing accuracy. However, the journey from theoretical model development to practical industrial implementation involves navigating numerous technical and operational challenges that significantly influence design decisions and system performance. In the following section, we summarize these implementation challenges and the strategies developed to address them, providing critical context for understanding both the achievements and limitations of our approach in real-world settings.

3.3.1. Data quality challenges

One of the fundamental challenges in developing accurate predictive models for PP dosing is managing data variability. As illustrated in Fig. 8, calibration procedures introduce distinct offsets and temporal patterns across data subsets. This heterogeneity required careful preprocessing strategies. The standardization process using StandardScaler proved crucial for ensuring comparable model performance across different volume ranges (0.3 ml vs. 1.2 ml), especially given the significant difference in absolute measurement values. Measurement precision presented another critical challenge in our experimental design. The accuracy of our models directly depends on the quality of the input data, making the precision of volume measurements a foundational element of our system. To address this challenge, we employed a Wipotec SL-M 250/300 high-precision scale with exceptional measurement specifications (resolution: 0.002 g, linearity: ±0.001 g, repeatability: 0.0005 g). To maintain this precision throughout our experiments, we implemented a rigorous quality assurance protocol following GAMP guidelines, including a three-point calibration procedure (0 g, 50 g, 100 g) before each experimental session. This calibration process ensured that measurement errors remained within the tolerance limits. The implementation of this quality control procedure was essential for establishing a reliable foundation for our AI models, as even minor measurement inconsistencies could potentially propagate through the system and significantly impact the effectiveness of our predictive algorithms. Our experimental setup also incorporated specific hardware elements to minimize external noise influence. Notably, the use of a high-precision weighing scale equipped with the Active Vibration Compensation (AVC) sensor developed by Wipotec significantly reduced interference from environmental vibrations, enhancing measurement consistency. For this technology to function properly, both the weighing cell and AVC sensor were rigidly connected on a collective base plate with sufficient rigidity. This mandatory precondition ensures

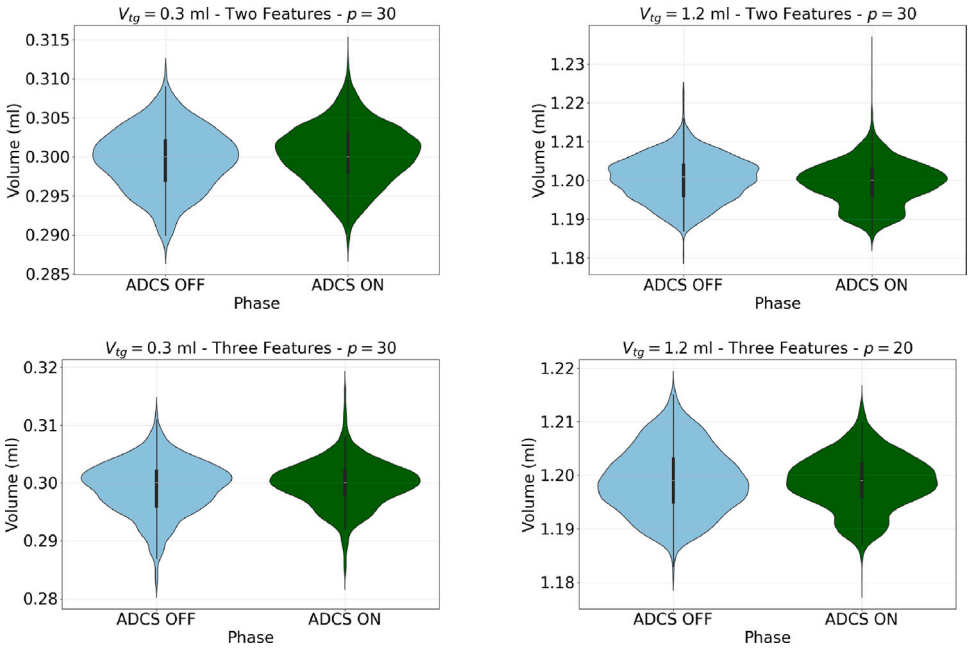

Fig. 16. Error distribution visualization comparing ADCS OFF and ADCS ON phases for Pre-trained approaches.

both systems experience identical vibration interference, allowing the AVC to effectively compensate for disturbances caused by immediate surrounding area movements (like machine frame vibrations). In our experimental setup, we ensured that the entire structure — comprising the base plate, AVC sensor, weighing cells, load carriers, and maximum weights — maintained a first natural frequency above 350 Hz in all directions, as required by the manufacturer for optimal performance. This hardware-level noise reduction complemented our software preprocessing techniques, creating a more robust foundation for the subsequent modeling stages.

3.3.2. Training parameter optimization challenges

The determination of optimal training parameters represented a significant challenge, requiring extensive empirical testing within practical constraints. The non-standard 85%–15% split ratio for training-validation partitioning emerged from a series of preliminary experiments specifically designed to address the limited dataset size typical in industrial pharmaceutical settings. Traditional splits (such as 70%–30% or 80%–20%) proved suboptimal in our context, providing either insufficient training examples or inadequate validation coverage. Through sequential testing of various configurations, we determined that an 85%–15% ratio offered the best balance for our specific application. Similarly, batch size selection involved significant empirical testing. For subset-specific training (Section 3.2.1), smaller batch sizes (8) demonstrated better convergence properties, while the complete dataset training (Section 3.2.3) benefited from larger batches (32). These findings emerged through empirical testing of several values within practical industrial constraints. The final parameters represent the best configurations identified within our experimental scope, acknowledging that theoretical optimization might yield incremental improvements.

The early stopping configuration presented another optimization challenge. Initial experiments with standard parameters (5–10 epochs patience) frequently resulted in premature training termination before the model could adequately capture temporal patterns. Conversely, excessively patient configurations led to overfitting and unnecessary computational expense. Our final approach with a 100-epoch warm-up period and 20-epoch patience parameter emerged as an empirical optimum after observing convergence patterns across multiple training runs. These settings effectively balanced learning capacity with practical training duration constraints.

This work addresses the dosing precision improvement blending theoretical and methodological investigation with a practical validation in an industrial setting. Through optimizing key architectural elements of LSTM and GRU networks and implementing targeted parameter selection, we demonstrate a positive and solid applicability to a complex industrial case, and across different volume ranges. The parameter configurations presented throughout this work deliver tangible improvements in real operating conditions while maintaining a limited computational effort for wide applicability.

3.3.3. Computational constraints

The industrial context imposed specific timing requirements that influenced our architectural decisions. As observed in our Online Training approach, the computational costs increased substantially with larger training windows (from approximately 10 s with TW = 100 to over 20 s with TW = 300). This performance consideration directly impacted the practical utility of certain configurations in real-time industrial settings, where cycle times constrain the available processing window.

Architecture selection involved careful consideration of both accuracy and computational efficiency. While LSTM and GRU networks demonstrated comparable prediction accuracy, the reduced parameter count of GRU models provided computational advantages critical for industrial deployment, particularly in scenarios requiring faster inference times. This trade-off analysis extended to neuron count determination, where we observed pronounced diminishing returns beyond 6 neurons (as shown in Figs. 11, 14, and 15). This finding guided our decision to limit model complexity, optimizing the balance between accuracy and computational efficiency.

4. Limitations & future work

While the results presented demonstrate significant improvements in peristaltic pump dosing accuracy, a detailed evaluation requires acknowledging the limitations of our current approach and identifying opportunities for further advancement.

The Online Training method, while effective, requires substantial computational resources, with execution times of approximately 10 s per cycle that may challenge integration in high-throughput industrial environments. However, this approach offers significant flexibility, as it can adapt to different setups, volumes, and dispensed products

without requiring pre-existing datasets specific to each scenario. Conversely, the Pre-trained approach, though computationally efficient, demonstrates reduced adaptability to significant changes in operational conditions, with effectiveness heavily dependent on the representativeness of historical training data for the specific setup being used. Pre-trained models cannot be assumed to perform optimally in new conditions without appropriate data collection and training for those specific scenarios. From a methodological perspective, our investigation focused on two specific volumes (0.3 ml and 1.2 ml), and generalization to volumes outside this range would require additional validation. Additionally, while our black-box approach was designed to be device-agnostic, all tests were conducted using a single peristaltic pump model, and validation across different systems would be necessary to confirm the broad applicability of our methods. These limitations, while not diminishing the significance of our findings, provide important considerations for researchers and practitioners seeking to apply these techniques in different contexts or scale them to production environments.

Looking ahead, this study suggests avenues for future research, particularly in the exploration of hybrid techniques: combining the strengths of AI and traditional statistical methods, investigating how to extract the maximum performance using them in conjunction and blending the respective outcomes based also on the required filling volume. Another promising direction for future work involves the integration of additional real-time sensor feedback to further enhance dosing accuracy. By incorporating complementary measurements such as tube pressure, temperature, or flow dynamics through strategically positioned sensors, the predictive models could potentially capture more nuanced aspects of the physical dispensing process. These multimodal sensing approaches could be particularly valuable for addressing the challenges encountered in micro-volume dispensing, where even minor variations can significantly impact dosing accuracy representing an important advancement toward adaptive control systems capable of maintaining optimal precision across ever-wider ranges of operating conditions.

5. Conclusions

This study presents an exploration of LSTM and GRU architectures to predict peristaltic pump (PP) dosages and enhance their accuracy through both an iterative Online Training and a Pre-trained approach. Improvements in dosing accuracy are very interesting for pharmaceutical manufacturing, which is the background of this work investigation, as well as for other contexts. The Online Training method yielded favorable results, particularly notable with a training window of 100 filling cycles and for dosages below 1.0 ml, for which the precision improvement reaches up to 38.4% in $RMSE_{PP}$ (Root Mean Square Error between actual and target dosages) and 31.6% in STD (Standard Deviation of actual dosages). For dosages exceeding 1.0 ml, the technique demonstrated a 23.2% gain in $RMSE_{PP}$. The execution time on our reference platform was approximately 10 s per cycle, which is compatible with the considered industrial application and easily acceleratable through faster GPU hardware. The Pre-trained approach, employing a three-features strategy, showed promising results, achieving statistically significant $RMSE_{PP}$ improvement of 9.2% for 0.3 ml dosages and both 4.6% STD and 13.8% $RMSE_{PP}$ gains for 1.2 ml dosages, with execution times below 1 s. Experimental results demonstrated that while both LSTM and GRU neural network architectures obtained similar predictive accuracy, GRU offered faster training times due to its lower number of parameters. Overall, these findings highlight that the optimal compensation approach varies depending on the dosage volume: for smaller volumes, AI-based methods outperform statistical approaches, while for larger volumes, some statistical models can be a viable alternative. Pre-trained models also demonstrate potential value when sufficient historical data is available.

CRediT authorship contribution statement

Davide Privitera: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. **Stefano Bellissima:** Investigation, Methodology, Software, Validation, Data curation. **Sandro Bartolini:** Conceptualization, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank Pharma Integration for providing with the opportunity to use the necessary equipment to conduct the experiments.

Data availability

Data will be made available on request.

References

Azhahudurai, K., & Veeramanikandan, D. V. (2023). Time series forecasting of air pollutant PM2.5 using transformer architecture. *International Journal of Science and Research (IJSR)*, URL: https://api.semanticscholar.org/CorpusID:266332467.

Banu, K. A., Aparna, T., Tajudeen, M. M., Rajchakit, G., & Huang, T. (2025). H∞ control for fractional order neural networks with uncertainties subject to deception attacks via improved memory-event-triggered scheme and its application. *Neural Networks*, *184*, Article 107092. http://dx.doi.org/10.1016/j.neunet.2024.107092, URL: https://www.sciencedirect.com/science/article/pii/S0893608024010219.

Bayat, S., & Işık, G. (2023). Assessing the efficacy of LSTM, transformer, and RNN architectures in text summarization. *International Conference on Applied Engineering and Natural Sciences*, *1*(1), 813–820. http://dx.doi.org/10.59287/icaens.1099, URL: https://as-proceeding.com/index.php/icaens/article/view/1099.

Buestán-Andrade, P.-A., Santos, M., Sierra-García, J.-E., & Pazmiño-Piedra, J.-P. (2023). Comparison of LSTM, GRU and transformer neural network architecture for prediction of wind turbine variables. In P. García Bringas, H. Pérez García, F. J. Martínez de Pisón, F. Martínez Álvarez, A. Troncoso Lora, Á. Herrero, J. L. Calvo Rolle, H. Quintián, & E. Corchado (Eds.), *18th international conference on soft computing models in industrial and environmental applications* SOCO 2023, (pp. 334–343). Cham: Springer Nature Switzerland.

Cao, L., Pan, Y., Liang, H., & Huang, T. (2023). Observer-based dynamic event-triggered control for multiagent systems with time-varying delay. *IEEE Transactions on Cybernetics*, *53*(5), 3376–3387. http://dx.doi.org/10.1109/TCYB.2022.3226873.

Chen, L., Zhu, Y., & Ahn, C. K. (2021). Adaptive neural network-based observer design for switched systems with quantized measurements. *IEEE Transactions on Neural Networks and Learning Systems*, 1–14. http://dx.doi.org/10.1109/TNNLS.2021.3131412.

Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder–decoder for statistical machine translation. In A. Moschitti, B. Pang, & W. Daelemans (Eds.), *Proceedings of the 2014 conference on empirical methods in natural language processing* EMNLP, (pp. 1724–1734). Doha, Qatar: Association for Computational Linguistics, http://dx.doi.org/10.3115/v1/D14-1179, URL: https://aclanthology.org/D14-1179.

Chollet, F., et al. (2015). Keras. https://keras.io. (Accessed 02 August 2024).

Cozzio, M., Melis, A., La Fauci, G., Guaraldi, P., Caputo, R., Lioi, F., Cellini, G. S., Santilli, G., Scarlattei, D., Siravo, P., Zuccheri, P., Ziglio, A., & Montalti, M. (2023). Vial sharing of high-cost drugs to decrease leftovers and costs: A retrospective observational study on Patisiran Administration in Bologna, Italy. *Healthcare*, *11*(7), http://dx.doi.org/10.3390/healthcare11071013.

European Union (2022). The rules governing medicinal products in the European union volume 4 EU guidelines for good manufacturing practice for medicinal products for human and veterinary use - annex 1 manufacture of sterile medicinal products. https://health.ec.europa.eu/medicinal-products/eudralex/eudralex-volume-4_en#part. (Accessed 02 August 2024).

Farid, S. S., Baron, M., Stamatis, C., Nie, W., & Coffman, J. (2020). Benchmarking biopharmaceutical process development and manufacturing cost contributions to R&D. *mAbs*, *12*(1), Article 1754999. http://dx.doi.org/10.1080/19420862.2020.1754999, PMID: 32449439.

Ferretti, P., Pagliari, C., Montalti, A., & Liverani, A. (2023). Design and development of a peristaltic pump for constant flow applications. *Frontiers in Mechanical Engineering*, *9*, http://dx.doi.org/10.3389/fmech.2023.1207464.

Francis, J., & Bian, L. (2019). Deep learning for distortion prediction in laser-based additive manufacturing using big data. *Manufacturing Letters*, *20*, 10–14. http://dx.doi.org/10.1016/j.mfglet.2019.02.001, URL: https://www.sciencedirect.com/science/article/pii/S221384631830172X.

Gasoto, S., Schneider, B., & Setti, J. (2022). Study of the pulse of peristaltic pumps for use in 3D extrusion bioprinting. *ACS Omega*, *7*, http://dx.doi.org/10.1021/acsomega.1c07093.

Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2017). LSTM: A search space odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, *28*(10), 2222–2232. http://dx.doi.org/10.1109/TNNLS.2016.2582924.

Harrou, F., Zeroual, A., Hittawe, M. M., & Sun, Y. (2022). Chapter 6 - recurrent and convolutional neural networks for traffic management. In F. Harrou, A. Zeroual, M. M. Hittawe, & Y. Sun (Eds.), *Road traffic modeling and management* (pp. 197–246). Elsevier, http://dx.doi.org/10.1016/B978-0-12-823432-7.00011-2, URL: https://www.sciencedirect.com/science/article/pii/B9780128234327000112.

Hittawe, M. M., Harrou, F., Sun, Y., & Knio, O. (2024). Stacked transformer models for enhanced wind speed prediction in the Red Sea. In *2024 IEEE 22nd international conference on industrial informatics* (pp. 1–7). http://dx.doi.org/10.1109/INDIN58382.2024.10774454.

Hittawe, M. M., Langodan, S., Beya, O., Hoteit, I., & Knio, O. (2022). Efficient SST prediction in the red sea using hybrid deep learning-based approach. In *2022 IEEE 20th international conference on industrial informatics* (pp. 107–117). http://dx.doi.org/10.1109/INDIN51773.2022.9976090.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. http://dx.doi.org/10.1162/neco.1997.9.8.1735.

Hu, C., Ou, T., Chang, H., Zhu, Y., & Zhu, L. (2020). Deep GRU neural network prediction and feedforward compensation for precision multiaxis motion control systems. *IEEE/ASME Transactions on Mechatronics*, *25*(3), 1377–1388. http://dx.doi.org/10.1109/TMECH.2020.2975343.

Jeppesen, J. (2009). MC100 pump control module user's manual. https://www.wmfts.com/globalassets/literature/m-flexicon-mc100-profibus-en.pdf. (Accessed 02 August 2024).

Jørgensen, F. (2008). PD12 OEM operators manual. https://archive.org/details/manualzilla-id-6886200. (Accessed 02 August 2024).

Klespitz, J., & Kovács, L. (2014). Peristaltic pumps — A review on working and control possibilities. In *2014 IEEE 12th international symposium on applied machine intelligence and informatics* (pp. 191–194). http://dx.doi.org/10.1109/SAMI.2014.6822404.

Lambert, P., & Joergensen, F. (2008). Accurate dispensing of biopharmaceuticals. *World Pumps*, *2008*, 22–24. http://dx.doi.org/10.1016/S0262-1762(08)70099-3.

Natof, T., & Pellegrini, M. V. (2025). *Food and drug administration recalls*. Treasure Island (FL): StatPearls Publishing LLC, Copyright © 2025.

Nuijten, M. (2022). Pricing Zolgensma – the world's most expensive drug. *Journal of Market Access & Health Policy*, *10*(1), Article 2022353. http://dx.doi.org/10.1080/20016689.2021.2022353.

Patel, R., Vhora, A., Jain, D., Patel, R., Khunt, D., Patel, R., Dyawanapelly, S., & Junnuthula, V. (2024). A retrospective regulatory analysis of FDA recalls carried out by pharmaceutical companies from 2012 to 2023. *Drug Discovery Today*, *29*(6), Article 103993. http://dx.doi.org/10.1016/j.drudis.2024.103993.

Peccerillo, B., Mannino, M., Mondelli, A., & Bartolini, S. (2022). A survey on hardware accelerators: Taxonomy, trends, challenges, and perspectives. *Journal of Systems Architecture*, *129*, Article 102561. http://dx.doi.org/10.1016/j.sysarc.2022.102561, URL: https://www.sciencedirect.com/science/article/pii/S1383762122001138.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Privitera, D., Bellissima, S., & Bartolini, S. (2023). Adaptive dosing control system through ARIMA model for peristaltic pumps. *IEEE Access*, *11*, 99558–99572. http://dx.doi.org/10.1109/ACCESS.2023.3314379.

Rajchakit, G., Banu, K. A., Aparna, T., & and, C. P. L. (2025). Event-triggered secure control for Markov jump neural networks with time-varying delays and subject to cyber-attacks via state estimation fuzzy approach. *International Journal of Systems Science*, *56*(2), 211–226. http://dx.doi.org/10.1080/00207721.2024.2390694, arXiv:https://doi.org/10.1080/00207721.2024.2390694.

Schäfer, M., Müller, T., & Fernandez, D. H. (2025). Deep learning for stock performance prediction: A sharpe ratio-optimized approach. *Asia Pacific Economic and Management Review*, URL: https://api.semanticscholar.org/CorpusID:277301173.

Shieu, W., Lamar, D., Stauch, O. B., & Maa, Y.-F. (2016). Filling of high-concentration monoclonal antibody formulations: Investigating underlying mechanisms that affect precision of low-volume fill by peristaltic pump. *PDA Journal of Pharmaceutical Science and Technology*, *70*(2), 143–156. http://dx.doi.org/10.5731/pdajpst.2015.005926, arXiv:https://journal.pda.org/content/70/2/143.full.pdf. URL: https://journal.pda.org/content/70/2/143.

Tariq, R. A., Vashisht, R., Sinha, A., & Scherbak, Y. (2023). *Medication dispensing errors and prevention*. StatPearls Publishing, Treasure Island (FL), URL: http://europepmc.org/books/NBK519065.

U. S. Food and Drug Administration (2015). Allowable excess volume and labeled vial fill size in injectable drug and biological products: Guidance for industry. https://tinyurl.com/fda-allowable-excess-volume. (Accessed 17 September 2024).

Wang, X., Bi, Q., Zhu, L., & Ding, H. (2018). Improved forecasting compensatory control to guarantee the remaining wall thickness for pocket milling of a large thin-walled part. *International Journal of Advanced Manufacturing Technology*, *94*, http://dx.doi.org/10.1007/s00170-016-9785-8.

Wang, Y., Zhang, H., Han, Z., & Ni, X. (2021). Optimization design of centrifugal pump flow control system based on adaptive control. *Processes*, *9*(9), http://dx.doi.org/10.3390/pr9091538, URL: https://www.mdpi.com/2227-9717/9/9/1538.

Wei, X., Wang, G., Schmalz, B., Hagan, D. F. T., & Duan, Z. (2023). Evaluation of transformer model and self-attention mechanism in the yangtze river basin runoff prediction. *Journal of Hydrology Regional Studies*, *47*, Article 101438. http://dx.doi.org/10.1016/j.ejrh.2023.101438, URL: https://www.sciencedirect.com/science/article/pii/S2214581823001258.

Wu, Q., Ma, L., Yu, P., Fu, R., & Feng, Z. (2024). Long-short term memory networks aided fault detection of power facilities. *Intelligent Systems with Applications*, *22*, Article 200395. http://dx.doi.org/10.1016/j.iswa.2024.200395, URL: https://www.sciencedirect.com/science/article/pii/S266730532400070X.

Wu, S., & Ni, J. (1989). Precision machining without precise machinery. *CIRP Annals*, *38*(1), 533–536. http://dx.doi.org/10.1016/S0007-8506(07)62762-0, URL: https://www.sciencedirect.com/science/article/pii/S0007850607627620.

Yan, L., & Huang, W. (2009). An IC yield enhancement approach by ARMA modeling and dynamic process control. *International Journal of Advanced Manufacturing Technology*, *42*(7), 749–756. http://dx.doi.org/10.1007/s00170-008-1631-1.