Real-time prediction of Atrial Fibrillation using Meta-learning

M. Moazeni[1 [0000−0002−4151−2446]], S. Mildiner Moraga[1 [0000−0001−6282−7394]], M. Wösten[2,3 []], Paul Elbers[4 [0000−0003−0447−6893]] L. W. van Laake[2 [0000−0002−5027−0661]], F. W. Asselbergs[5,6,7 [0000−0002−1692−8669]], and E. Aarts[1 [0000−0002−2432−7564]]

[1] Department of Methodology and Statistics, Utrecht University, Utrecht, the Netherlands m.moazeni@uu.nl
[2] Department of Cardiology, University Medical Centre Utrecht, Utrecht University, Utrecht, the Netherlands
[3] Computational Imaging Group for MRI Therapy & Diagnostics, University Medical Centre Utrecht
[4] Intensive care, Amsterdam UMC, Vrije Universiteit, Amsterdam
[5] Institute of Cardiovascular Science, Faculty of Population Health Sciences, University College London, London, UK
[6] Health Data Research UK and Institute of Health Informatics, University College London, London, UK
[7] Department of Cardiology, Amsterdam Cardiovascular Sciences, Amsterdam University Medical Centre, University of Amsterdam, Amsterdam, The Netherlands.

Abstract. Atrial Fibrillation (AF), a heart rhythm disorder causing an irregular and rapid heart rate, affects up to 44% of ICU patients. Early prediction of AF in the ICU can enable timely interventions, reducing negative outcomes like prolonged length of stay (LOS) and increased mortality. We developed a continuous risk prediction model to detect AF onset using numerical ICU data. Data from ICU patients with a minimum stay of 6 hours were collected from AmsterdamUMCdb, Europe's first open-access ICU database. This dataset included numerical values such as heart rate and blood pressure, along with medication and fluid administration records. A Long Short-Term Memory (LSTM) deep learning model was developed, employing a Model-agnostic Meta-Learning (MAML) approach to ensure generalizability across different patients. To address dataset imbalance, additional importance weights were applied in the binary cross-entropy log loss function. The model was then tested on an imbalanced test set that represented the real-world ratio of AF to non-AF patients. The model achieved AUC scores of 0.92 and an accuracy of 0.88. SHAP values were used to understand the model's decision-making. The LSTM model within the MAML framework effectively recognized non-AF conditions but was limited in identifying AF instances.

Keywords: Meta-learning · LSTM · Real-time prediction

1 Introduction

New-onset atrial fibrillation (AF), often occurring as a common complication in critical illness, is a condition well-known to intensive care unit (ICU) doctors, affecting up to 44% of ICU admitted patients [4,8,12]. Detecting AF onset in ICU can enable timely interventions, thereby decreasing negative hospital outcomes during a patient's stay [10].

Identifying patients at risk of AF in the ICU is not commonly practiced, and predicting it in a timely manner can be difficult [10, 17]. Previous research has explored AF detection models through two approaches: patient-wise prediction, which assigns a single AF risk score to each patient for their entire ICU stay, and real-time prediction, which generates a risk score for each measurement recorded during the patient's time in the ICU.

For the patient-wise predictions, studies relied on using electrocardiogram (ECG) signals for AF detection reporting sensitivity and accuracy ranging between 40% to 100% and 70% to 98%, respectively [2,3,9]. However, these models focus on diagnosing AF a short time before its occurrence, and they rely on expert interpretation which restricts the size and diversity of patient cohorts, limiting their applicability as real-time decision support systems in the practice. Furthermore, due to their high measurement frequency ECG data are not routinely stored for ICU patients which impacts their accessibility for analysis.

In contrast, a number of studies have emphasized the value of routinely collected healthcare numerical data from the ICU, such as vital signs, medications, and fluid balance, in predicting AF , showing that ICU routinely collected numerical data having valuable information for AF detection [7, 11, 11, 14, 17]. Reported precision and recall in these studies ranges from 7% to 74% and 15% to 75%, respectively, showing significant variability in the effectiveness of the models, which indicates that there is room for improvement in the methodologies employed.

While patient-wise models offer valuable insights into the likelihood of a patient developing AF, they do not indicate the timing of its onset. Consequently, some studies have adopted the strategy of aggregating measurements within specific time windows and making predictions for each window. For example by trying different time windows ranging from 12 to 24h, Verhaege et.al developed a semi real-time model for AF detection [17]. However, in the ICU where timely intervention is critical, enabling the model to provide real-time AF risk scores can help the clinicians to make more informed decisions. This has the potential of improving patient outcomes by allowing earlier interventions.

In this study, our objective is to develop a real-time prediction model that classifies each measurement as AF (atrial fibrillation) or non-AF during a patient's ICU stay, aiming to improve performance measures for both groups. The model will utilize numerical values collected during the ICU stay to continuously provide a risk score for each measurement. By incorporating meta-learning, the model can dynamically adapt to new and diverse patient data, improving its generalization across different scenarios. Additionally, zero-shot learning equips the model with the capability to make accurate predictions without providing

additional data for fine-tuning, thereby enhancing its robustness. To accomplish this, we integrate zero-shot meta-learning with Long Short-Term Memory (LSTM) networks, leveraging the strengths of both approaches to enhance the model's predictive accuracy and adaptability [15]. By incorporating zero-shot meta-learning with LSTM, we aim to capture temporal patterns in patients' time-series data and accommodate inter-patient variability effectively.

2 Materials and Methods

Our model fitting and evaluation of the proposed approach utilize the Amsterdam University Medical Centers Database (AmsterdamUMCdb) ICU database (v1.02). This database, endorsed by the European Society of Intensive Care Medicine (ESICM), stands as the inaugural freely accessible European intensive care unit (ICU) database [16]. The database contains information from a 32-bed combined surgical-medical academic intensive care unit (ICU) and a 12-bed intermediate care facility [16]. For patients who experienced atrial fibrillation (AF) during ICU stay, it was determined based on the first time AF was recorded by nursing staff after at least one sinus rhythm registration. The data processing and analysis pipeline followed is visualized in Figure 1, which depicts the stages from raw data acquisition to the final model evaluation. In the upcoming paragraphs each of these stages are briefly discussed.

To ensure a sufficient data set for training and testing the model, patients are included regardless of their total length of stay (LOS). The last 2 hours before the event—either AF onset for AF patients or the last recorded time for non-AF patients—are excluded from the analysis. This period, termed the buffer window, prevents the model from accessing the final 2 hours before the event, thus providing a potential window for clinical interventions aimed at preventing AF. Within the analyzed time frame, the final 24 hours are identified as the "Pre-AF window." This window is crucial for detecting atrial fibrillation (AF), highlighting the period where the model focuses on identifying the onset of AF (see Figure 1). For patients who do not stay for at least 24 hours, the Pre-AF window covers the duration from admission to their LOS, and these hours are used to label AF occurrence. For non-AF patients we marked their last measurement as the endpoint. The methodology also designates all the measurements prior to Pre-AF window as No-AF window, assigning 0 for the labels.

To ensure consistency in the frequency of consecutive measurements, we standardized the frequency by averaging the measurements within each hour to achieve one measurement per hour (i.e., interval harmonization). Additionally, we introduced a new feature into the model to indicate the number of measurements that have been averaged.

Variables with more than 90% missing data across the dataset were filled in using mean imputation, while those with less than 90% were imputed using Multiple Imputation by Chained Equations (MICE) (full details are available in Supplementary). The baseline characteristics of the patient population are summarized in Table 1.

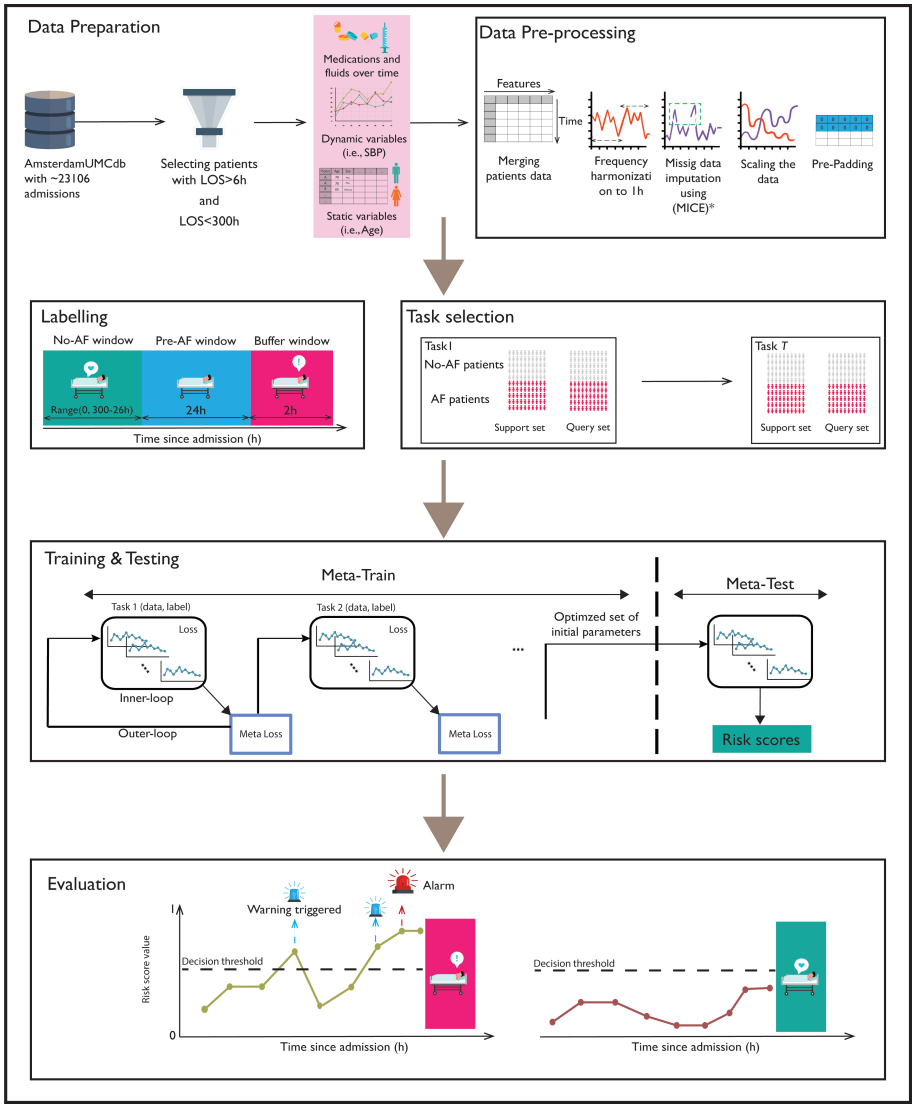

Fig. 1. Model development overview. The process begins with data preparation, where patient data is filtered based on length of stay, and relevant dynamic and static variables are extracted. In the data pre-processing step, features are merged, measurement frequencies harmonized, missing data imputed, and the data scaled and pre-padded. During labeling, AF labels are assigned to the 24-hour period prior to AF onset, excluding a 2-hour buffer window. Task selection involves grouping patients into tasks for training, with separate support and query sets. The training and testing phase employs a meta-learning approach with inner-loop and outer-loop processes to optimize model parameters. Finally, during evaluation, the model's risk scores are monitored over time, triggering alarms if thresholds are exceeded, indicating a high risk of AF.

	AF patients	No-AF patients
Cohort size	1307	16904
Age (group)	70.2 (69-75)	61.0 (55.0-75.0)
Gender (Male)	807 (61.8%)	10940 (64.2%)
BMI	26.5 (23.5-27.4)	25.4 (23.4- 27.1)
Weight	77.4 (65.0-85.0)	77.7 (65.0-85.0)
Mortality	669 (51.1%)	4467 (26.2%)
Sepsis	298 (22.8%)	1787 (10.5%)
Had a cardiology surgery	335 (25.6%)	5039 (30.0%)
SOFA (first 24h)	8.4 (6-11)	5.2 (3-8)
APACHE II (first 24h)	22 (17-26)	16.6 (12-20)
ICU LOS (days)	14.3 (3.8-18.9)	3.0 (1.3- 2.4)
ICU survival	46.4%	73.4%

Table 1. Baseline characteristics of patients with and without AF used for training and evaluation

2.1 Model Development

Our study employs the Model-Agnostic Meta-Learning (MAML) approach to enhance the robustness and generalization of a model dedicated to atrial fibrillation (AF) detection [6]. Meta-learning aims to broaden the model's adaptability to various tasks, thereby enhancing its effectiveness in identifying AF from heterogeneous patient data. A "task" in our study is defined as a distinct learning challenge presented to the model, formulated to differentiate between patients with AF and those without (non-AF patients) based on collected measurements throughout the ICU stay. Each task consists of a support set and a query set:

- **Support set** (S): The support set includes labeled data for initial model fitting to the task, comprising full sequence data from (N_1) patients with AF and (N_2) non-AF patients. This set trains the model to identify distinguishing features and patterns between different patients.
- **Query set** (Q): The query set assesses the model's fitting, containing data from (M_1) AF patients and (M_2) non-AF patients, all previously unseen in the support set. This set tests the model's learning and generalization capability.

The loss for task T_i is obtained as :

$$\mathcal{L}_{T_i}(f_\theta) = \frac{1}{|Q_i|} \sum_{(x,y)\in Q_i} \text{BinaryCrossEntropy}(f_{\theta'}(x), y) \tag{1}$$

where θ represents the initial parameters of the model and θ' represents the updated parameters after training on the support set S_i.

At the core of our methodology lies the Long Short-Term Memory (LSTM) neural network, serving as the base learner. Base learner in this context refers to the specific model or algorithm that is directly trained on a given task's data to make predictions (a.k.a *inner loop updates*). The LSTM base learner in our study

is adept at processing and making predictions based on full sequence patient data, by leveraging its ability to remember information from admission point to AF onset. This step, focusing on inner loop updates, aims to find optimize parameters in identifying both AF and non-AF measurements within the task data utilized for model training. The update rule for the inner loop is given by:

$$\theta' = \theta - \alpha \nabla_\theta \mathcal{L}_{S_i}(f_\theta) \tag{2}$$

Here, α is the inner loop learning rate, and $\mathcal{L}_{S_i}(f_\theta)$ is the loss on the support set.

Building upon the fitted model from the inner loop updates, the MAML framework is designed to enable rapid model adjustments to new patient data, instead of specializing in a single, specific task. This is achieved by monitoring the loss function values on the query set, referred to as *outer loop updates*. The primary goal of MAML in this context is to refine the base learner LSTM's initial parameters to ensures that these parameters are sufficiently generalized across a variety of patient profiles. The MAML objective function can be represented as:

$$\min_\theta \sum_i \mathcal{L}_{T_i}(f_{\theta'}) \tag{3}$$

and the MAML update rule is:

$$\theta \leftarrow \theta - \beta \nabla_\theta \sum_i \mathcal{L}_{T_i}(f_{\theta'}) \tag{4}$$

Where β is the outer loop learning rate.The complete loss function for the MAML in our study, incorporating the inner and outer updates, can be expressed as:

$$\mathcal{L}_{\text{meta}}(\theta) = \sum_i \mathcal{L}_{Q_i}(f_{\theta - \alpha \nabla_\theta \mathcal{L}_{S_i}(f_\theta)}) \tag{5}$$

For the training, the Adam optimizer is selected for its ability to adjust learning rates adaptively.

2.2 Hyperparameter optimization & Cross validation

To enhance the training of our model, we undertook hyperparameter optimization. The parameters N1 (representing number of non-AF patients) and N2 (representing number of AF patients were adjusted between 100 to 300, in increments of 100. Additionally, for the weighting factor in the loss function, we experimented with values from 1 to 10, increasing by steps of 2. To further ensure robust and reliable results, we employed 5-fold cross-validation.

2.3 Feature selection and evaluation

The models were constructed using Python, version 3.9.16. For the meta-learning algorithm, the learn2learn package was utilized alongside PyTorch's LSTM module (PyTorch version 2.1) [1]. The output from the models were transformed

using a sigmoid function to yield probability scores that range from 0 to 1. A threshold of 0.5 was then applied to the AF risk scores to trigger alerts. The first alert is treated as a precautionary warning, so it is not categorized as either a true positive or a false positive. Should another alarm occur within the next hour, it is recognized as an alarm, and any further alarms are treated similarly. This precautionary mechanism is employed once in the 24 hours leading up to the onset of AF and again from the time of admission up until the 24 hours before onset, aiming to enhance understanding in both periods.

The performance of the model was assessed using various metrics, including precision, recall, F1 score, and Area Under the Curve (AUC), drawn from a 20% held-out test set of patients, preserving the class imbalance present in real-world data. We believe it is crucial to evaluate the model's effectiveness in a real-world scenario characterized by class imbalance; therefore, the test set was deliberately kept imbalanced to reflect this.

Precision measures the accuracy of the positive alerts generated by the model, indicating the proportion of true positive alerts out of all positive alerts (true positives and false positives). Recall (or sensitivity) assesses the model's ability to identify all actual cases of AF, calculated as the proportion of true positive alerts out of the total actual AF measurements. F1 Score is a harmonic mean of precision and recall, providing a single metric to assess the balance between them, with a higher score indicating better performance. Area Under the Curve (AUC) relates to the ROC curve and measures the model's ability to distinguish between the classes (AF and non-AF measurements), with a higher value indicating better discrimination.

We adopted the same variable selection procedure as outlined by Verhaeghe et al. (2023). Initially, 282 unique potential variables were selected for inclusion in the AmsterdamUMCdb. Variables with fewer than 250 occurrences or with over 99% missing data were excluded, resulting in 194 variables. Additionally, we included the number of measurements within an hour before harmonizing frequency, the time difference between these measurements, and the number of fluid measurements within an hour before harmonization.

2.4 Risk scores for Test patients

Given the focus on clinical application, grasping how the proposed model performs in this environment is crucial. Consequently, the model was assessed using a separate set of test patients (20% of the total cohort), monitoring them from ICU admission up to AF occurrence for AF patients or up to 300 hours for non-AF patients. To analyze how the model predicts risk over time, we examined the risk scores of patients during their final 48 hours of stay. These scores were averaged for all patients at each time point and smoothed to illustrate the overall trend. This approach allows for determining whether the model steadily or slightly increases the risk for AF patients nearing the onset of AF, for instance, demonstrating stable or slightly increasing alarms over time when compared to non-AF patients.

SHAP analysis To illuminate the mechanisms and outcomes of the final models, SHAP analysis was performed to enhance explainability. Each SHAP value is linked to a specific feature within a data sample, clearly indicating the impact that feature has on the model's prediction in comparison to the model's average prediction across all data samples. By utilizing SHAP analysis on patient risk scores, our goal is to closely examine their progression. Through the evaluation of SHAP values, we obtain a deeper understanding of the role of individual features in the fluctuation of risk scores throughout the last 36 hours of stay. This approach allows us to discern how the importance of specific feature values in influencing the model's predictions shifts as the patient's condition changes.

3 Results

In the AmsterdamUMCdb dataset, out of 4,086 patients diagnosed with atrial fibrillation (AF) and 19,020 without AF, we included 1,307 AF patients and 16904 non-AF patients in our study. Patients were excluded based on LOS threshold explained in section 2 (see figure 1). Twenty percent of these patients were allocated to the test set, consisting of 261 AF patients and 3389 non-AF patients, while the remaining patients participated in the model training.

Table 2. Performance metrics for AF and Non-AF patients in different periods. A is the balanced test set (containing 261 AF and non-AF patients) and B is an unbalanced test set (containing 261 AF patients and 3380 non-AF patients). P and N are pre-AF and non-AF windows, respectively.

Metric	**Cohort**	**P-period**		**N-period**	
		AF pat.	**Non-AF pat.**	**AF pat.**	**Non-AF pat.**
Precision	A	0.75 [0.03]	0.93 [0.02]	NR	NR
	B	0.29 [0.05]	0.97 [0.01]	NR	NR
Recall	A	0.75 [0.07]	0.83 [0.06]	0.83 [0.04]	0.93 [0.02]
	B	0.78 [0.04]	0.79 [0.06]	0.94 [0.01]	0.94 [0.01]
F1	A	0.75 [0.03]	0.87 [0.00]	NR	NR
	B	0.42 [0.04]	0.87 [0.04]	NR	NR
N.Measurements	A	5764	6634	72653	72461
	B	6502	62968	82351	946961
% of patients detected correctly	A	0.56 [0.03]	0.89 [0.03]	0.85 [0.01]	0.95 [0.01]
	B	0.52 [0.02]	0.83 [0.01]	0.94 [0.01]	0.94 [0.01]
Average LOS (h)	A	17.2	18.2	88.0	75.8
	B	16.9	17.7	86.9	77.3

3.1 Model outcome

Performance metrics for AF and Non-AF patients in different periods. A is the balanced test set (containing 261 AF and non-AF patients) and B is an unbal-

anced test set (containing 261 AF patients and 3380 non-AF patients). P and N are pre-AF and non-AF windows, respectively.

The model test evaluation metrics for the balanced and unbalanced test sets are shown on Table 2 for AF and Non-AF patients in the two periods (i.e., pre-AF and no-AF). Overall, the model achieved an Area Under the Curve (AUC) of 92% in the unbalanced test set. In the unbalanced test set, the model's precision was 29% for AF patients in the pre-AF period. This indicates that in 29% of cases where the model sounded the alarm, an actual AF event was occurring. This level of precision enabled the model to correctly identify 78% of AF events, which translates to 137 patients.

For non-AF instances in pre-AF window, the precision was 97%, reflecting the model's heightened accuracy in identifying true non-AF instances, therefore minimizing unnecessary alarms. Looking at recall in the No-AF period, for both AF and Non-AF was 96%, indicating high reliability in not raising false alarms when patients are not yet in the critical Pre-AF period. Additionally, the model achieved an overall accuracy of 91%, reflecting its stable performance in different clinical situations where the frequency of AF can vary.

3.2 Risk scores over time

Figure 2 provides a visual representation of the risk scores predicted by the model for both AF and non-AF patients over the last 48 hours of stay. The average risk scores of patients are plotted to illustrate changes over time, with the shaded area indicating the standard deviation around these averages.

For AF patients, the average predicted line (indicated by the solid blue line) shows an upward trend as the time approaches the onset of AF. This suggests that the model is increasingly confident in predicting an AF event as it gets closer to occurring. The shaded area around the the line indicates the variability in the predictions, with a narrower band suggesting more consistent predictions by the model. In contrast, the non-AF patient measurements, represented by the black lines, remain relatively stable and lower throughout the last 48 hours of stay, indicating that the model consistently assigns lower risk scores to this group.

3.3 Alarms distribution

The percentage of alarms depicts the distribution of alarms over the last 36 hours for both AF patients and non-AF patients, representing the percentage of patients who had an alarm at each respective hour (Figure 3).

For AF patients, the graph displays a fairly consistent distribution of alarms across the 36-hour period leading up to the onset of AF. The percentages fluctuate slightly but generally remain above 40%. This consistent level of alarms may indicate that the model is sensitive to changes that occur in the physiological parameters that it monitors, suggesting a persistent risk of AF throughout this time frame.

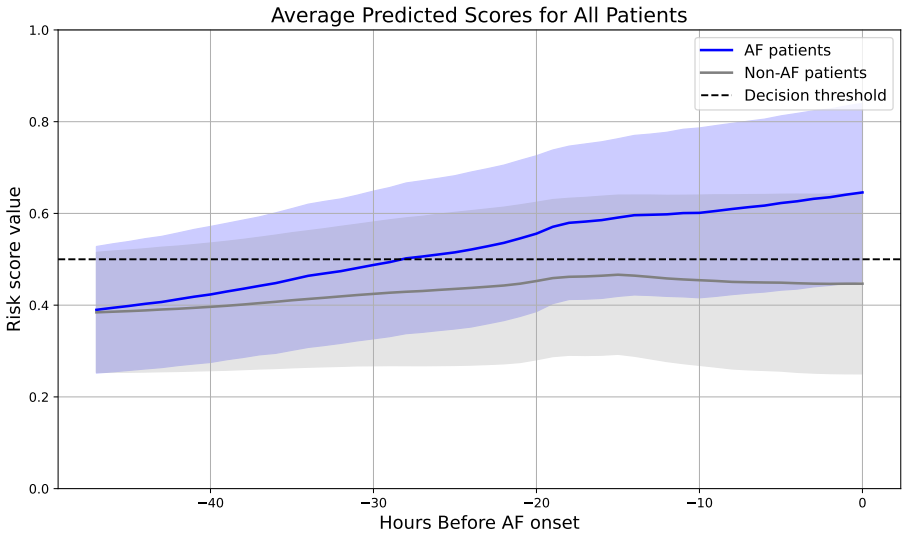

Fig. 2. Average risk scores for patients with atrial fibrillation (AF) are depicted by the blue line, while scores for non-AF patients are shown with the grey line. The shaded blue and grey areas represent the standard deviation from the average for AF and non-AF patients, respectively. The dashed line indicates the decision threshold of 0.5, used to differentiate between alarm and no alarm measurements.

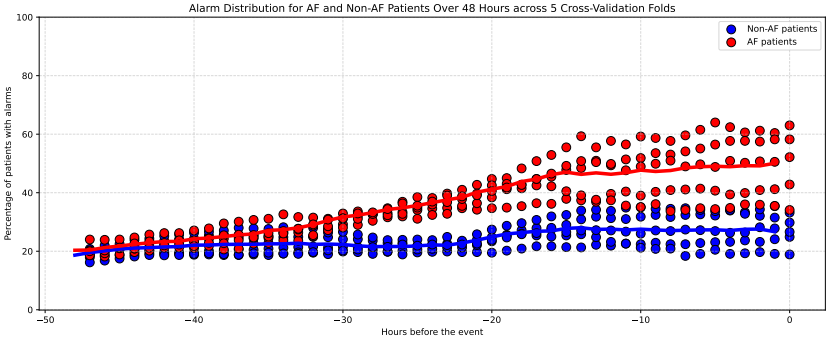

Fig. 3. The distribution of alarms over the last 48 hours is shown with red bars for AF patients and blue bars for non-AF patients. For AF patients, the data represents the period prior to an AF event, whereas for non-AF patients, it covers their entire 48-hour stay

In the case of non-AF patients, the line chart shows a significantly different pattern. The percentage of alarms is consistently lower compared to AF patients, mostly remaining under 30%. This suggests that the model is less likely to predict an AF event for these patients, which is expected as they do not actually experience an AF onset.

3.4 Shapley Analysis

The SHAP (SHapley Additive exPlanations) summary plot provides insight into the feature importance determined by the predictive model (figure 4). Key clinical features such as invasive blood pressure, Lipase in blood or plasma and heart rate emerged as strong influencers on the model's risk score prediction. The spread of SHAP values for each feature indicates the variability of their impact on the predictions, with Age being notably significant. This underscores the nuanced interplay between patient-specific clinical parameters and their contribution to the model's ability to forecast atrial fibrillation events accurately.

4 Discussion

We conducted a retrospective study to evaluate the performance of a real-time prediction model designed to early predict the onset of AF in ICU patients. The model was tested on the AmsterdamUMCdb dataset in two settings: balanced and unbalanced test datasets. Models was capable of achieving high AUC score of 92% and improved precision score of 29% for AF patients and 89% for non-AF patients in comparison to comparable datasets [14,17] while predicting AF onset in real-time for each hour of stay in ICU. In addition, the predicted risk scores for AF patients were generally higher compared to those for non-AF patients. An analysis of the most critical features, based on SHAP values, identified invasive blood pressure, Lipase in blood or serum, and the heart rate as key determinants.

The study has several limitations. Firstly, it was not possible to retrospectively identify patients who developed AF after discharge. For these patients, the model might have generated alarms close to their discharge, which would have been recorded as false positives. Additionally, the diagnosis of AF relies on nursing charts as electrocardiograms were not available in the AmsterdamUMCdb. This approach assumes that nurses' recordings are accurate and timely to avoid data leakage and these recordings have been deemed sufficient [5, 17]. Though machine learning models that use ECG waveforms can improve AF detection and the speed of diagnosis, incorporating ECG data could enhance performance. Despite this, our method effectively provides risk predictions using only routinely collected data. It's also important to mention that ECG signals might miss paroxysmal AF and require expert annotations, which could limit their practical application. (References to be added later). Second, in the unbalanced test set, although the model achieves a higher AUC and precision compared to earlier studies, the current level of precision may still result in alarm fatigue. This may be attributed to the low incidence rate of the primary outcome. One possible

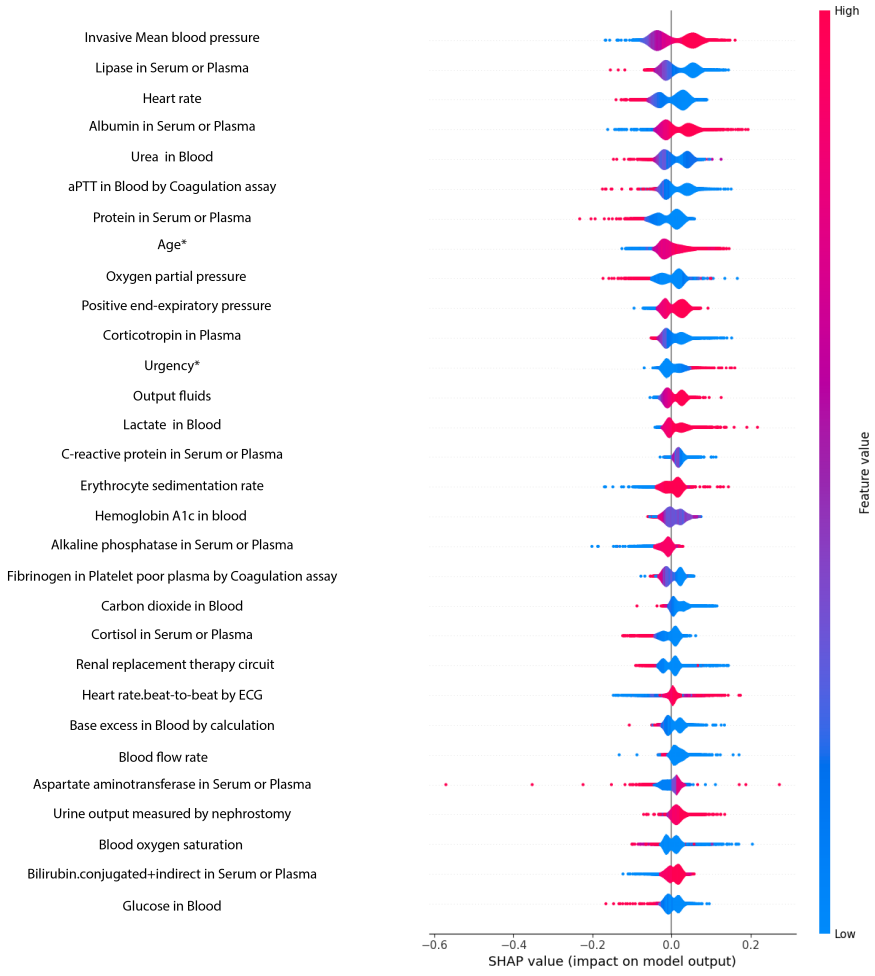

Fig. 4. Shapley analysis of the model.

solution is by adjusting the decision threshold for each patient separately to manage alarm fatigue more closely [13].

During the harmonization process, we averaged the frequency of measurements for each feature per patient to a uniform hourly rate, while also recording the original number of measurements. However, this approach might lead to the loss of valuable information. The varying number of measurements is a direct result of clinical decisions in the ICU, over which we had no control. As such, the impact of these varying frequencies on the model's output remains unknown. This variability introduces the potential for bias due to the intensity of monitoring—referred to as bias by indication. Specifically, changes in patient predicted risk scores patterns, often in anticipation of AF, become key features in the model. Consequently, the model might underperform in situations where healthcare providers have not already recognized signs of AF, undermining the effectiveness of an the model designed to alert to such conditions. This issue is particularly pertinent for prescribed medications, which are typically ordered in response to specific concerns about a patient's condition.

We assumed that early signs of AF could occur within 24 hours prior to its onset. This window was necessary for training and evaluating the algorithm. However, we recognize that using shorter windows of 24 hours presents a more challenging task due to increased dataset imbalance and reduced data availability for the algorithm to learn the progression of AF over time. In addition, we set a minimum LOS requirement of 6 hours for patients so each patient has enough measurements to train and test the model on. This requirement poses a problem in real-life scenario, as it excludes patients who may develop AF early during their ICU stay. To address this issue, alternative methods should be explored to include these patients in the training and testing phases of the algorithm. One such alternative is to simulate data. By generating synthetic data for patients with shorter LOS, we can augment the dataset, ensuring the algorithm is exposed to a broader range of LOS's.

5 Conclusion

Our research concentrated on developing a real-time atrial fibrillation prediction model through a meta-learning approach. This model generates risk scores ranging from 0 to 1 and issues an alarm when these scores surpass the decision boundary. We trained and assessed the model using the AmsterdamUMCdb dataset, the first openly accessible European dataset. The model achieved a high AUC of 92%, demonstrating its effectiveness in distinguishing between negative and positive classes. Although the precision of this model has improved compared to previous studies, further enhancements are necessary.

Acknowledgements

We thank Jarne Verhaeghe and Thomas De Corte for making publicly accessible the code to pre-process AmsterdamUMCdb. This work made use of the

Dutch national e-infrastructure Research Cloud with the support of the SURF Cooperative using grant no. EINF-7919, which is (partly) financed by the Dutch Research Council (NWO). All funding sources had no role in the design of this study, nor did they have any role during its execution, analyses, interpretation of the data, and decision to submit results.

Declaration of interest

The authors declare no conflicts of interest.

Open Practices and data Sharing

The data that support the findings of this study are available at `https://amsterdammedicaldatascience.nl`. The code behind this study has been made publicly available at GitHub and can be accessed at `https://github.com/Mehranmzn/AF_MAML`.

References

1. Arnold, S.M.R., Mahajan, P., Datta, D., Bunner, I., Saitas Zarkias, K.: learn2learn: A Library for Meta-Learning Research (Aug 2020), `https://github.com/learnables/learn2learn`, original-date: 2019-08-08T18:22:41Z
2. Attia, Z.I., Noseworthy, P.A., Lopez-Jimenez, F., Asirvatham, S.J., Deshmukh, A.J., Gersh, B.J., Carter, R.E., Yao, X., Rabinstein, A.A., Erickson, B.J., Kapa, S., Friedman, P.A.: An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. The Lancet **394**(10201), 861–867 (Sep 2019). https://doi.org/10.1016/S0140-6736(19)31721-0, `https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(19)31721-0/abstract`, publisher: Elsevier
3. Boon, K.H., Khalil-Hani, M., Malarvili, M.: Paroxysmal atrial fibrillation prediction based on HRV analysis and non-dominated sorting genetic algorithm III. Computer Methods and Programs in Biomedicine **153**, 171–184 (Jan 2018). https://doi.org/10.1016/j.cmpb.2017.10.012, `https://www.sciencedirect.com/science/article/pii/S0169260717301815`
4. Bosch, N.A., Cimini, J., Walkey, A.J.: Atrial Fibrillation in the ICU. Chest **154**(6), 1424–1434 (Dec 2018). https://doi.org/10.1016/j.chest.2018.03.040, `https://www.sciencedirect.com/science/article/pii/S0012369218305452`
5. Ding, E.Y., Albuquerque, D., Winter, M., Binici, S., Piche, J., Bashar, S.K., Chon, K., Walkey, A.J., McManus, D.D.: Novel Method of Atrial Fibrillation Case Identification and Burden Estimation Using the MIMIC-III Electronic Health Data Set. Journal of Intensive Care Medicine **34**(10), 851–857 (Oct 2019). https://doi.org/10.1177/0885066619866172, `https://doi.org/10.1177/0885066619866172`, publisher: SAGE Publications Inc STM
6. Finn, C., Abbeel, P., Levine, S.: Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks (Jul 2017). https://doi.org/10.48550/arXiv.1703.03400, `http://arxiv.org/abs/1703.03400`, arXiv:1703.03400

7. Jia, Z., Shi, Y., Hu, J.: Personalized Neural Network for Patient-Specific Health Monitoring in IoT: A Metalearning Approach. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems **41**(12), 5394–5407 (Dec 2022). https://doi.org/10.1109/TCAD.2022.3162182, conference Name: IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems
8. Johnston, B.W., Chean, C.S., Duarte, R., Hill, R., Blackwood, B., McAuley, D.F., Welters, I.D.: Management of new onset atrial fibrillation in critically unwell adult patients: a systematic review and narrative synthesis. British Journal of Anaesthesia **128**(5), 759–771 (May 2022). https://doi.org/10.1016/j.bja.2021.11.016, https://www.sciencedirect.com/science/article/pii/S0007091221007443
9. Kim, J., Sangjun, O., Kim, Y., Lee, M.: Convolutional Neural Network with Biologically Inspired Retinal Structure. Procedia Computer Science **88**, 145–154 (Jan 2016). https://doi.org/10.1016/j.procs.2016.07.418, https://www.sciencedirect.com/science/article/pii/S187705091631674X
10. Klein Klouwenberg, P.M.C., Frencken, J.F., Kuipers, S., Ong, D.S.Y., Peelen, L.M., van Vught, L.A., Schultz, M.J., van der Poll, T., Bonten, M.J., Cremer, O.L.: Incidence, Predictors, and Outcomes of New-Onset Atrial Fibrillation in Critically Ill Patients with Sepsis. A Cohort Study. American Journal of Respiratory and Critical Care Medicine **195**(2), 205–211 (Jan 2017). https://doi.org/10.1164/rccm.201603-0618OC, https://www.atsjournals.org/doi/10.1164/rccm.201603-0618OC, publisher: American Thoracic Society - AJRCCM
11. McMillan, S., Rubinfeld, I., Syed, Z.: Predicting atrial fibrillation from intensive care unit numeric data. In: 2012 Computing in Cardiology. pp. 213–216 (Sep 2012), iSSN: 2325-8853
12. Meierhenrich, R., Steinhilber, E., Eggermann, C., Weiss, M., Voglic, S., Bögelein, D., Gauss, A., Georgieff, M., Stahl, W.: Incidence and prognostic impact of new-onset atrial fibrillation in patients with septic shock: a prospective observational study. Critical Care **14**(3), R108 (Jun 2010). https://doi.org/10.1186/cc9057, https://doi.org/10.1186/cc9057
13. Movahedi, A., Sadooghiasl, A., Ahmadi, F., Vaismoradi, M.: Smart care for dealing with nurses' alarm fatigue in the intensive care unit. Journal of Nursing Scholarship **55**(4), 825–833 (2023). https://doi.org/10.1111/jnu.12870, https://onlinelibrary.wiley.com/doi/abs/10.1111/jnu.12870, _eprint: https://sigmapubs.onlinelibrary.wiley.com/doi/pdf/10.1111/jnu.12870
14. Ortega-Martorell, S., Pieroni, M., Johnston, B.W., Olier, I., Welters, I.D.: Development of a Risk Prediction Model for New Episodes of Atrial Fibrillation in Medical-Surgical Critically Ill Patients Using the AmsterdamUMCdb. Frontiers in Cardiovascular Medicine **9** (2022), https://www.frontiersin.org/articles/10.3389/fcvm.2022.897709
15. Rafiei, A., Moore, R., Jahromi, S., Hajati, F., Kamaleswaran, R.: Meta-learning in healthcare: A survey (Aug 2023). https://doi.org/10.48550/arXiv.2308.02877, http://arxiv.org/abs/2308.02877, arXiv:2308.02877
16. Thoral, P.J., Peppink, J.M., Driessen, R.H., Sijbrands, E.J.G., Kompanje, E.J.O., Kaplan, L., Bailey, H., Kesecioglu, J., Cecconi, M., Churpek, M., Clermont, G., van der Schaar, M., Ercole, A., Girbes, A.R.J., Elbers, P.W.G.: Sharing ICU Patient Data Responsibly Under the Society of Critical Care Medicine/European Society of Intensive Care Medicine Joint Data Science Collaboration: The Amsterdam University Medical Centers Database (AmsterdamUMCdb) Example*. Critical Care Medicine **49**(6), e563–e577 (Jun

2021). https://doi.org/10.1097/CCM.0000000000004916, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8132908/

17. Verhaeghe, J., De Corte, T., Sauer, C.M., Hendriks, T., Thijssens, O.W.M., Ongenae, F., Elbers, P., De Waele, J., Van Hoecke, S.: Generalizable calibrated machine learning models for real-time atrial fibrillation risk prediction in ICU patients. International Journal of Medical Informatics **175**, 105086 (Jul 2023). https://doi.org/10.1016/j.ijmedinf.2023.105086, https://www.sciencedirect.com/science/article/pii/S1386505623001041