Machine Learning Approaches in Atrial Fibrillation Detection

Deividas Butkus, Jolita Bernatavičienė

Vilnius University, Institute of Data Science and Digital Technologies
Akademijos str. 4, Vilnius
deividas.butkus@mif.stud.vu.lt

Abstract. Atrial fibrillation (AF) characterized by rapid and irregular electrical activity in the atria represents a prevalent form of cardiac arrhythmia that significantly challenges healthcare systems due to its links to heightened mortality and morbidity rates. Early detection of AF is critical for accurate and effective management and treatment. In response to this pressing need, numerous researchers have used machine learning (ML) to enhance the precision and efficiency of AF detection. By analyzing available datasets, signal lengths, preprocessing techniques, and a diverse array of ML approaches, this paper aims to cover methodologies of AF detection using electrocardiogram (ECG) data and ML.

Keywords: atrial fibrillation, machine learning, ECG, artificial intelligence, deep learning

1 Introduction

Atrial fibrillation (AF) is the most common cardiac arrhythmia seen in clinical settings leading to serious health problems such as stroke, heart failure, and increased mortality rates. The electrocardiogram (ECG) provides a graphical representation of the heart's electrical activity over time and is the standard diagnostic tool for detecting AF. The normal electrocardiogram in sinus rhythm depicted in Figure 1 comprises a P wave, a QRS complex, and a T wave. The QRS complex is often, but not always, three separate waves: the Q wave, the R wave, and the S wave. When AF is present, ECG usually comprises the absence of a P wave (while the QRS complex and T waves remain present), and an irregular pattern of R-waves.

The diagnosis of AF based on ECG signals requires the expertise of a trained specialist, typically a doctor, making it a time and resource-intensive process. The need for human interpretation poses challenges, including potential delays in diagnosis and limitations in scalability. Given these

DOI: https://doi.org/10.15388/LMITT.2024.2

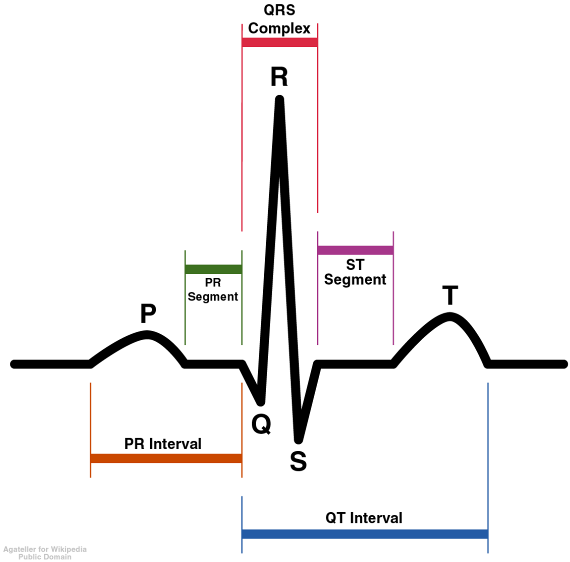

Figure 1. Schematic diagram of normal sinus rhythm for a human heart as seen on ECG [1]

challenges and the evolving landscape of healthcare technology, there is an increasing recognition of the potential role of machine learning algorithms in AF detection. Understanding nuances of the development of machine learning (ML) algorithms capable of identifying AF from ECG data with high accuracy is crucial for the advancement of automated diagnostic tools. This paper explores various machine learning approaches applied to ECG data for the detection of AF, aiming to systematically present advancements in AF detection in the ML field. Through an examination of different methodologies, from data preprocessing to model selection, this analysis seeks to review the most common strategies for this vital diagnostic task.

2 Literature review

2.1 Databases

Many recent publications in the field of ML for AF detection use public datasets provided by PhysioNet [2], a data repository for biomedical research. Based on the research by [3], the MIT-BIH Atrial Fibrillation Database (AFDB) [4] and the Computing in Cardiology (CinC) Challenge 2017 Database [5] are the two most popular databases for AF detection, used in around 50 studies out of 147 reviewed, and both provided by PhysioNet. The AFDB database consists of 25 long-term ECG recordings of human subjects with AF, while the CinC Challenge 2017 contains a training set with 8528 single lead ECG recordings (normal (5076), AF (758), other (2415), and noisy (279)) lasting 9–60 s and a test set with 3658 ECG recordings of similar lengths that have been retained as a hidden test set. This training dataset is unbalanced and skewed towards the normal sinus rhythm class. Other databases, such as MIT-BIH Arrhythmia (MITDB) [6], and MIT-BIH Normal Sinus Rhythm (NSRDB) [7] are also often used. Based on the recent review of ML in AF detection, 10 out of 14 papers reviewed also used the AFDB Database, showing the continued relevance and popularity of this database in the field [8].

2.2 Signal length

The determination of an ECG signal length for the detection of AF using ML can depend on several factors, such as the objectives of the study, and the practical considerations of the machine learning algorithm. The authors of [9] tested various signal lengths for paroxysmal atrial fibrillation classification and got the best results with a 4 s window using the Second-Order System (SOS) algorithm. Another study by [10] employed both 2 s and 5 s windows using Convolutional Neural Network (CNN) and found that the 2 s segments achieved a higher specificity while 5 s segments showed a slightly better overall accuracy and sensitivity. However, study by [11] also used CNN and varying windows of 9–60 seconds and concluded that it was difficult to distinguish AF from other rhythms on small signal segments. Longer signal of 31 heartbeats was also backed up by [12] where the combination of CNN and Recurrent-Neural Networks (RNN) achieved specificity and sensitivity of 98.96% and 86.04% on unseen data. This variation in segment window and ML methods underscores the absence of a universal standard in the selection process. Optimal signal duration should satisfy the prerequisites of machine learning algorithms while simultaneously capture the dynamic nature of AF to facilitate precise classification.

2.3 ECG preprossessing techniques and features extraction

Morphological characteristics of the ECG are often derived and widely used in ML-based AF detection. One part of them is called time-domain transformations and includes RR interval, Heart Rate Variability (HRV), and

P-wave. Another group of transformations work on the frequency domain to detect high vs. low-frequency segments of the ECG and requires the use of Fourier Transform (FT) or Wavelet Transform (WT). These two domains, separately or together, are used widely in the research [13], [14, 15], [16]. In recent years, models such as CNN and RNN have been employed that directly process raw ECG signals, simplifying the detection process by eliminating the need for complex preprocessing steps [17], [18], [19]. This shift from complex preprocessing techniques to methods requiring minimal or no ECG preparation highlights an increasing trend in developing more efficient and accurate machine learning-based approaches for AF detection.

2.4 Machine learning algorithms

Machine learning algorithms for AF detection have demonstrated significant diversity and innovation, integrating traditional machine learning approaches with sophisticated deep learning models to enhance diagnostic accuracy. Deep learning methods, particularly CNN and Long Short-Term Memory (LSTM) networks, have shown to surpass traditional classifiers like Multilayer Perceptrons (MLP) and logistic regression in effectively processing ECG signals for AF detection, indicating a shift towards more complex models for better performance [20].

Further advancements include the use of CNNs in both single-channel and innovative two-channel models. A two-channel CNN model, for instance, uses one channel to identify where to look for the detection of AF in the ECG, while the other performs the actual detection [19]. Additionally, CNNs have been employed directly on raw ECG waveforms, bypassing traditional feature extraction processes [21], and in combination with RNN for extracting high-level features from RR intervals [12]. The study by [17] combined CNN and bagged tree ensemble to classify a filtered ECG signal - if the confidence of the classifier reaches a certain threshold, CNN is used, otherwise 43 PQRS features are used in a bagged tree ensemble.

Ensemble models and decision trees have also been instrumental in AF detection, utilizing a combination of expert features, signal processing methods, and learned features. These models benefit from the ensemble strategy by integrating multiple classifiers to improve prediction outcomes, evidenced by the use of bagged tree ensembles, gradient-boosted tree, and random forest classifiers based on hand-crafted and selected features for reliable AF detection [22], [23], [24].

Recent developments in the field have introduced innovative methods to enhance the detection and classification of AF. One such method utilizes a deep residualdense network based on a bidirectional recurrent neural network (Bi-RNN). This approach combines one-dimensional dense residual networks with Bi-RNNs and attention mechanisms to enable end-to-end feature learning from ECG signals, simplifying labour-intensive feature extraction steps [25]. Another approach merges the strengths of multilayer CNNs and RNNs with LSTM capabilities into a singular classifier. First, the multilayer CNN is utilized for extracting high-level features from the raw input sequence and then the RNN structure known as LSTM is used for processing the sequential features extracted by CNN. Lastly, the logistic classifier provides the posterior probability of the input sequence containing AF with notable sensitivity, specificity, and accuracy rates [26].

In conclusion, the automated AF detection landscape shows significant variation across ML algorithms without a single best method for AF detection. This diversity is highlighted in Table 1, which collates and compares different databases, signal lengths, preprocessing techniques, and machine learning algorithms with their corresponding performance metrics.

2.5 Evaluation metrics

The evaluation of ML algorithms in the context of AF detection is crucial to ensure their reliability and effectiveness in clinical applications. Traditional metrics for assessing the accuracy of detection methods in medicine include sensitivity, specificity, positive predictive value, and accuracy [29], [30]. These metrics serve to gauge:

Table 1. Summary of AF detection studies with percentage evaluation scores.

Author	Dataset	Signal length	Features	ML algorithm	Eval. metric	Score
Andersen et al (2017) [15]	AFDB	60, 100, and 300 beats	Time-domain, frequency-domain features	Support Vector Machine	Se, Sp	96.81%, 96.20%
Zabihi et al (2017) [24]	CinC Challenge 2017	5s segments with 4s overlap	Time-domain, frequency-domain, nonlinear features, meta-level features, morphological features	Random forest	F1	84%

Author	Dataset	Signal length	Features	ML algorithm	Eval. metric	Score
Hong et al (2017) [22]	CinC Challenge 2017	20 s	Expert features, center-wave features, and DNN features	Decision trees	F1	85%
Kamaleswaran et al (2018) [21]	CinC Challenge 2017	9s to 61s	Raw ECG data at various sampling frequencies	CNN	F1	82%
Kropf et al (2018) [23]	CinC Challenge 2017	Not specified	time-domain, frequency-domain, and morphological features	Gradient boosted tree	F1	84%
Plesinger et al (2018) [17]	CinC Challenge 2017	6 s	PQRS features or a 9-times filtered ECG signal	Bagged tree or CNN	F1	82%
Andersen et al (2017) [15]	AFDB	60, 100, and 300 beats	Time-domain, frequency-domain features	Support Vector Machine	Se, Sp	96.81%, 96.20%
Mousavi et al (2019) [19]	AFDB	5 s	Raw ECG data	ECGNET	Se, Sp, Acc	99.53%, 99.26%, 99.40%
Laghari et al. (2023) [25]	CPSC2018 [27]	Not specified	Residual dense CNN and RNN features	Residual dense CNN and RNN	Se, Sp, Acc	93.09%, 98.71%, 97.72%
Kumar et al. (2023) [26]	CACHET-CADB [28], AFDB, NSRDB, MITDB	Interval of 30 RR	Multilayer CNN features	CNN and RNN	Se, Sp, Acc	96.06%, 98.29%, 97.04%

- Sensitivity (Se), also known as true positive rate (TPR) or recall, measures the proportion of signals correctly classified as AF versus the actual number of signals identified as AF.
- Specificity (Sp) measures the proportion of negative cases that are correctly classified as not AF versus all signals identified as not AF in the observed set.
- Positive Predictivity Value (PPV), also known as precision, is the proportion of signals correctly classified as AF versus the total number of AF in the set.

- Accuracy (Acc) measures the overall model's performance and is calculated as a ratio of correctly predicted observations (both AF and non-AF) to the total observations in the proportion.

However, the performance of algorithms submitted to the CinC Challenge 2017 was measured using the F1 score [5], a metric that provides a balanced view of the model's proficiency in accurately classifying both positive and negative instances. The F1 score is calculated as:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

An empirical study on the performance evaluation of AF detection highlighted the F1 score's suitability as the premier metric for assessing AF detection algorithms [31]. This research confirms the effectiveness of the F1 score through a comprehensive examination of various data sets, promoting the value of the F1 score in giving a detailed understanding of how well an algorithm can accurately detect true positives while effectively minimizing false positives and negatives. This robust methodological approach makes a strong case for using the F1 score to evaluate how well algorithms detect atrial fibrillation.

3 Conclusions

This paper systematically reviewed the application of ML algorithms in the detection of AF using ECG data. Through the examination of databases, signal lengths, preprocessing techniques, and ML algorithms, it has been shown that ML can significantly enhance the precision and efficiency of AF detection. The analysis revealed a preference for certain public datasets, such as those provided by PhysioNet, and highlighted the absence of a universal standard for optimal signal length in AF detection, which can vary based on the objectives of the study and the capabilities of the ML algorithm used.

Preprocessing techniques for ECG signal analysis have evolved, with recent trends showing a shift towards models that require minimal or no preprocessing, like CNNs and RNNs. These advancements suggest a move towards more direct analysis of raw ECG signals, simplifying the detection process. Furthermore, the exploration of various ML algorithms, including deep learning models like CNNs and LSTMs, demonstrated their advancement over traditional machine learning models due to their ability to process complex patterns within ECG signals more effectively. However, it should be also highlighted that no single best method for AF detection has emerged due to variations in ML methodologies and datasets.

The evaluation of the algorithms using metrics such as sensitivity, specificity, and the F1 score is important for determining their applicability in clinical settings. The findings support the use of the F1 score as a balanced measure of an algorithm's ability to accurately classify AF, which is vital for developing reliable diagnostic tools. However, even though the studies reviewed achieved high prediction scores, the results cannot be compared with each other because of the differences in the datasets and methodology used. Also, most of the studies evaluated the performance of ML algorithms using data from the same database used for training rather than unseen, real-life data. The research does not sufficiently clarify if the training and testing data consist of distinct patient groups. This distinction is critical because if a long ECG signal is segmented and portions of it are used for both training and testing, the algorithm's performance may appear artificially inflated, as it could learn and thus anticipate the heart's variability within an individual during training. It is imperative for future research to conduct cross-database testing, address data imbalance, and ensure that the training and testing datasets are truly independent to avoid inflated performance results.

The variance in signal lengths and preprocessing techniques, alongside the diverse array of machine learning algorithms, illustrates the complexity of achieving a standardized approach. Future studies should focus on refining ML models to improve their diagnostic accuracy on real-life data and establish benchmarks that enable the comparison of AF detection methods. Also, the next steps in ML for AF detection should strive for clinical validation. The ultimate goal is to integrate these ML algorithms seamlessly into clinical workflows, contributing to the early and precise detection of AF and improving patient outcomes. This endeavour requires a multi-faceted approach, combining the technical advancements in ML with rigorous clinical testing to establish the practical efficacy of these automated systems.

References

[1] "SinusRhythmLabels.png — wikipedia, the free encyclopedia," 2006. [Online; accessed 9-April-2024].

[2] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, “PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals,” *Circulation*, vol. 101, pp. e215–e220, June 2000.

[3] I. Olier, S. Ortega-Martorell, M. Pieroni, and G. Y. Lip, “How machine learning is impacting research in atrial fibrillation: implications for risk prediction and future management,” *Cardiovascular Research*, vol. 117, no. 7, pp. 1700–1717, 2021.

[4] G. Moody, “A new method for detecting atrial fibrillation using rr intervals,” *Proc. Comput. Cardiol.*, vol. 10, pp. 227–230, 1983.

[5] G. D. Clifford, C. Liu, B. Moody, H. L. Li-wei, I. Silva, Q. Li, A. Johnson, and R. G. Mark, “Af classification from a short single lead ecg recording: The physionet/computing in cardiology challenge 2017,” in *2017 Computing in Cardiology (CinC)*, pp. 1–4, IEEE, 2017.

[6] G. B. Moody and R. G. Mark, “The impact of the mit-bih arrhythmia database,” *IEEE engineering in medicine and biology magazine*, vol. 20, no. 3, pp. 45–50, 2001.

[7] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, “Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals,” *circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[8] C. Xie, Z. Wang, C. Yang, J. Liu, and H. Liang, “Machine learning for detecting atrial fibrillation from ecgs: Systematic review and meta-analysis,” *Reviews in Cardiovascular Medicine*, vol. 25, no. 1, p. 8, 2024.

[9] N. A. Abdul-Kadir, N. M. Safri, and M. A. Othman, “Effect of ecg episodes on parameters extraction for paroxysmal atrial fibrillation classification,” in *2014 IEEE Conference on Biomedical Engineering and Sciences (IECBES)*, pp. 874–877, IEEE, 2014.

[10] U. R. Acharya, H. Fujita, O. S. Lih, Y. Hagiwara, J. H. Tan, and M. Adam, “Automated detection of arrhythmias using different intervals of tachycardia ecg segments with convolutional neural network,” *Information sciences*, vol. 405, pp. 81–90, 2017.

[11] Z. Xiong, M. K. Stiles, and J. Zhao, “Robust ecg signal classification for detection of atrial fibrillation using a novel neural network. in 2017 computing in cardiology (cinc),” *IEEE. https://doi. org/10.22489/CinC*, vol. 138, 2017.

[12] R. S. Andersen, A. Peimankar, and S. Puthusserypady, “A deep learning approach for real-time detection of atrial fibrillation,” *Expert Systems with Applications*, vol. 115, pp. 465–473, 2019.

[13] V. Gliner and Y. Yaniv, “An svm approach for identifying atrial fibrillation,” *Physiological Measurement*, vol. 39, no. 9, p. 094007, 2018.

[14] N. Sadr, M. Jayawardhana, T. T. Pham, R. Tang, A. T. Balaei, and P. de Chazal, “A low-complexity algorithm for detection of atrial fibrillation using an ecg,” *Physiological measurement*, vol. 39, no. 6, p. 064003, 2018.

[15] R. S. Andersen, E. S. Poulsen, and S. Puthusserypady, “A novel approach for automatic detection of atrial fibrillation based on inter beat intervals and support vector machine,” in *2017 39th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, pp. 2039–2042, IEEE, 2017.

[16] R. Smisek, J. Hejc, M. Ronzhina, A. Nemcova, L. Marsanova, J. Kolarova, L. Smital, and M. Vitek, “Multi-stage svm approach for cardiac arrhythmias detection in short single-lead ecg recorded by a wearable device,” *Physiological measurement*, vol. 39, no. 9, p. 094003, 2018.

[17] F. Plesinger, P. Nejedly, I. Viscor, J. Halamek, and P. Jurak, "Parallel use of a convolutional neural network and bagged tree ensemble for the classification of holter ecg," *Physiological measurement*, vol. 39, no. 9, p. 094002, 2018.

[18] K.-S. Lee, S. Jung, Y. Gil, and H. S. Son, "Atrial fibrillation classification based on convolutional neural networks," *BMC medical informatics and decision making*, vol. 19, pp. 1–6, 2019.

[19] S. Mousavi, F. Afghah, A. Razi, and U. R. Acharya, "Ecgnet: Learning where to attend for detection of atrial fibrillation with deep visual attention," in *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, pp. 1–4, IEEE, 2019.

[20] S. Liaqat, K. Dashtipour, A. Zahid, K. Assaleh, K. Arshad, and N. Ramzan, "Detection of atrial fibrillation using a machine learning approach," *Information*, vol. 11, no. 12, p. 549, 2020.

[21] R. Kamaleswaran, R. Mahajan, and O. Akbilgic, "A robust deep convolutional neural network for the classification of abnormal cardiac rhythm using single lead electrocardiograms of variable length," *Physiological measurement*, vol. 39, no. 3, p. 035006, 2018.

[22] S. Hong, M. Wu, Y. Zhou, Q. Wang, J. Shang, H. Li, and J. Xie, "Encase: An ensemble classifier for ecg classification using expert features and deep neural networks," in *2017 Computing in cardiology (cinc)*, pp. 1–4, IEEE, 2017.

[23] M. Kropf, D. Hayn, D. Morris, A.-K. Radhakrishnan, E. Belyavskiy, A. Frydas, E. Pieske-Kraigher, B. Pieske, and G. Schreier, "Cardiac anomaly detection based on time and frequency domain features using tree-based classifiers," *Physiological measurement*, vol. 39, no. 11, p. 114001, 2018.

[24] M. Zabihi, A. B. Rad, A. K. Katsaggelos, S. Kiranyaz, S. Narkilahti, and M. Gabbouj, "Detection of atrial fibrillation in ecg hand-held devices using a random forest classifier," in *2017 Computing in Cardiology (CinC)*, pp. 1–4, IEEE, 2017.

[25] A. A. Laghari, Y. Sun, M. Alhussein, K. Aurangzeb, M. S. Anwar, and M. Rashid, "Deep residual-dense network based on bidirectional recurrent neural network for atrial fibrillation detection," *Scientific Reports*, vol. 13, no. 1, p. 15109, 2023.

[26] D. Kumar, S. Puthusserypady, H. Dominguez, K. Sharma, and J. E. Bardram, "An investigation of the contextual distribution of false positives in a deep learning-based atrial fibrillation detection algorithm," *Expert Systems with Applications*, vol. 211, p. 118540, 2023.

[27] F. Liu, C. Liu, L. Zhao, X. Zhang, X. Wu, X. Xu, Y. Liu, C. Ma, S. Wei, Z. He, *et al.*, "An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection," *Journal of Medical Imaging and Health Informatics*, vol. 8, no. 7, pp. 1368–1373, 2018.

[28] D. Kumar, S. Puthusserypady, H. Dominguez, K. Sharma, and J. E. Bardram, "Cachet-cadb: A contextualized ambulatory electrocardiography arrhythmia dataset," *Frontiers in Cardiovascular Medicine*, vol. 9, p. 893090, 2022.

[29] H. B. Wong and G. H. Lim, "Measures of diagnostic accuracy: sensitivity, specificity, ppv and npv," *Proceedings of Singapore healthcare*, vol. 20, no. 4, pp. 316–318, 2011.

[30] W. Zhu, N. Zeng, N. Wang, *et al.*, "Sensitivity, specificity, accuracy, associated confidence interval and roc analysis with practical sas implementations," *NESUG proceedings: health care and life sciences, Baltimore, Maryland*, vol. 19, p. 67, 2010.

[31] M. Gusev and M. Boshkovska, "Performance evaluation of atrial fibrillation detection," in *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pp. 342–347, IEEE, 2019.