Proceedings of the 3rd International Conference on Biological Engineering and Medical Science
DOI: 10.54254/2753-8818/23/20231053

Machine learning based techniques for ECG noise removal and feature extraction

Xiaohan Gu

Hong Kong Baptist University, Faculty of Science, 224 Waterloo Road, Kowloon Tsai, Hong Kong, HK, CHN

21253250@life.hkbu.edu.hk

Abstract. Machine learning (ML) is being applied to all aspects of life with the development of artificial intelligence (AI). This paper explores the application of machine learning technology in electrocardiogram (ECG) analysis to diagnose and classify a patient's current cardiac disease, predict possible future diseases, and provide a personalised treatment plan. However, several challenges have been highlighted. First, individual ECG signals display variability, causing concern about effective diagnosis based on changing ECG data. Second, different diseases can produce similar ECG results, requiring powerful classification algorithms to accurately classify diseases. Finally, using patient information to predict the probability of future heart attacks is critical to developing appropriate prevention and treatment strategies. Overcoming these challenges could revolutionise the field of cardiology. It could enable precise and proactive medical intervention. The study highlights the potential of machine learning to improve cardiovascular care and personalised medicine and emphasises the importance of addressing key challenges to maximise its impact in clinical practice.

Keywords: machine learning, electrocardiogram, healthcare.

1. Introduction

AI and machine learning are rapidly driving change across various industries, and healthcare is no exception. In recent years, major companies such as Amazon and Apple have been actively engaged in machine learning projects to provide more intelligent responses to people's needs, such as personalised product recommendations and advanced voice recognition. By embracing the era of big data, AI can enable more personalised and intelligent services, increasing productivity and efficiency. ECG analysis is one of the areas where machine learning is expected to transform healthcare. The ECG is a crucial biomedical signal used to diagnose a variety of heart conditions, providing invaluable clinical information to healthcare professionals.

However, there are still some challenges in applying machine learning technology to ECG analysis: first, the original ECG has a lot of noise, and it is difficult to remove these noises; second, because each person's ECG has certain differences, or two different heart diseases have similar ECG signals, which leads to the problem that the machine cannot be correctly identified; third, whether the machine can use suitable methods such as supervised learning and unsupervised learning to accurately predict the patient's future condition based on the patient's current health status, and give correct medical

DOI: 10.54254/2753-8818/23/20231053

advice. In order to better innovate the existing ECG signal analysis technology, it is necessary to use machine learning to refine and objectify the behavior, which is in line with the trend of the times.

The purpose of this paper is to show that it is possible to apply machine learning technology to the analysis of ECG signals, and to perform noise reduction based on the three specific noise sources: baseline wonder, power line interference and muscle contractions, and to select the Appropriate classification algorithms for feature classification and selection of appropriate machine learning methods for model construction are discussed from three perspectives. Hoping that this article can help to advance the field of ECG technology going forward.

The application of machine learning (ML) to electrocardiogram (ECG) analysis aims to achieve three key goals: accurate diagnosis of heart disease, identification of specific heart disease types for each patient, and predicting the likelihood of developing other medical conditions within a specified timeframe. By realizing these functions, this technology can significantly enhance existing ECG analysis methods, replacing subjective and error-prone manual approaches. Leveraging vast data, machine learning systems excel in recognizing intricate patterns and abnormalities in ECG signals, leading to improved accuracy in diagnosing heart diseases and enabling timely interventions. Moreover, personalized treatment solutions can be provided by analyzing individual ECG data and comparing it to historical cases. Predicting future health risks allows for proactive preventive measures. Embracing machine learning in ECG analysis revolutionizes cardiology, promoting objective and data-driven practices, thus empowering healthcare professionals and patients alike for better outcomes.

2. Signal processing and noise removal in ECG signals

The ability to remove noise from ECGs suggests that machine learning can be applied to ECG analysis. Based on the three main noise baseline wonder, power line interference and muscle contractions introduced earlier. The current technology has been able to denoise these three types of noise. To be clear, there are P wave, the QRS complex, and the T wave in the ECG.

First of all, the baseline wonder will produce certain noise interference. According to Censi et al [1], the baseline wonder will affect the P wave, thereby affecting the interpretation of ECG signals. Mneimneh et al [2]. An adaptive Kalman filter (KF) is proposed to remove baseline miracles in real time. This opens up a good idea for removing the baseline wonder. Compared with other methods, the KF method has the least distortion. Although the KF method of removing the baseline will fail in the case of high-frequency changes, this problem can be solved by increasing the window size of the temporal samples.

Among them, power line interference and muscle contraction will affect the entire ECG signal. So it is very important to remove them. Mahesh Chawan [3] designed digital finite impulse response (FIR) and other ripple notch filters, which can effectively eliminate power line interference, improve signal quality and facilitate accurate diagnosis. Not only power line interference can be resolved, but also the noise produced by muscle contraction can be eliminated. Sharma and Sidhu [4] showed that muscle contractions can be filtered using least mean square (LMS) and normalized LMS (NLMS) algorithms to remove noise and obtain a pure ECG signal.

After these three noises are removed, ECG signal preprocessing can be performed. The idea of applying machine learning techniques to ECGs is nonsense if the noise cannot be removed. In addition, Fourier transform and wavelet transform can also be used for signal processing to decompose the signal into components that appear at different scales. Therefore, machine learning can remove noise and provide a good basis for analyzing ECG.

3. Feature extraction for ECG signal classification

The ability to perform feature extraction is also important evidence supporting the implementation of machine learning techniques for ECG interpretation. As highlighted by Ramadevi et al [5], an efficient and selective feature set is crucial for feature dimensionality reduction. Regarding the classification of the disease, since there is no optimal classification rule for ECG classification, the classification is particularly difficult. If the feature dimension is too complex, its feasibility will be reduced. Due to

DOI: 10.54254/2753-8818/23/20231053

differences in heart and physiology, each individual may exhibit unique ECG features. Therefore, creating universal diagnostic criteria becomes complicated and requires advanced signal processing techniques to accommodate individual differences. If this difficulty can be resolved, machine learning for ECG analysis will become more practical.

In addition, certain different heart diseases may exhibit similar effects on healthy ECG signals, further complicating the diagnostic process. This phenomenon can lead to misdiagnosis and requires powerful machine learning algorithms to identify subtle differences. It is also one of the most significant challenges in detecting heart disease using ECG signals.

There is now a statistical technique for reducing the dimensionality of datasets called Principal Component Analysis (PCA). The process involves ignoring class labels, computing mean vectors for each dimension, generating covariance matrices, computing eigenvalues and eigenvectors, and selecting the most informative eigenvectors to transform the samples into a new subspace. The application of machine learning technology can simplify the high-dimensional feature space and contribute to the efficient and accurate analysis of ECG data. According to Gosling and Butch [6], PCA has hitherto been used to address signal processing problems, most importantly ECG data compression and clinically indicative issues related to the characterization and diagnosis of myocardial width, ventricular repolarization, and atrial fibrillation. That is to say, if we can rely on this technology to reduce the dimensionality of data sets, then using machine learning to parse ECG will become a reality.

A case study by Giuliani [7] demonstrates the effectiveness of PCA in reducing the dimensionality of features in animal behavior studies. In ECG analysis, segmentation methods divide the ECG signal into different component waveforms, reducing redundant information and simplifying the authentication process. Feature extraction is a key step in extracting data to effectively distinguish different users. During the enrollment phase, the extracted feature set can be stored on a smart card, or stored in a remote database as a template indexed by user identity information. Multiple registered ECG signals from the same user were averaged to create this template. Authentication is facilitated by hardware or software comparing a new input query to a template to determine if the user's biometrics match the template. After such a large amount of data is stored in an orderly manner, it can become a database, so as to classify and organize it, and also serve as the basis for subsequent machine learning.

4. Machine learning algorithms for ECG signal classification

Machine Learning Algorithms for ECG Signal Classification is the most critical step and the strongest evidence to prove that the technology can be realized. In ECG signal processing, the initial step often involves supervised classification, where known abnormal ECG signals are used to train the model. This helps the algorithm learn to differentiate between normal and abnormal patterns. Once the model is trained, it can then be applied to new, unseen ECG signals to classify them as either normal or abnormal. In some cases, unsupervised learning may also be employed to explore the data further and identify any hidden patterns or structures that could aid in the interpretation of ECG signals. Unsupervised techniques can be particularly useful when the available labeled data is limited, and the model needs to discover additional insights from the data. In conclusion, machine learning plays a crucial role in interpreting ECG signals by identifying specific patterns and classifying them as normal or abnormal. Supervised learning utilizes labeled data to train the model, while unsupervised learning helps uncover hidden patterns and structures in unlabeled data. Both approaches contribute to a comprehensive understanding of ECG signals and support accurate diagnosis and classification of heart conditions. Therefore, this technology can be used for disease prediction, so as to achieve better medical treatment in advance. It can be seen that ECG analysis can be done with machine learning technology.

5. Possible technology gap

The proposed framework for ECG analysis utilizing machine learning undoubtedly represents an important step forward; however, it is critical to acknowledge and address existing technical gaps and limitations. There are several areas that require further attention and improvement to ensure the effective, efficient and ethical use of this technology.

First, noise removal in ECG signals remains a key technical challenge. The accuracy and reliability of machine learning algorithms depend heavily on the quality of the input data. Therefore, developing more powerful noise removal techniques and preprocessing methods is crucial to improve the overall performance of ECG analysis systems.

Second, practical considerations such as time and cost constraints pose realistic obstacles to the widespread adoption of this technology. For example, the computational complexity of some machine learning models can hinder real-time analysis in resource-constrained environments. Efforts should be made to optimize algorithms and deploy hardware that can efficiently process ECG data without compromising accuracy. Additionally, financial constraints may limit access to advanced ECG analysis tools in poorer communities. Addressing these cost issues and working to make these technologies affordable and accessible to all are critical to equitable healthcare progress.

In addition, user privacy and data confidentiality issues require high attention. Because machine learning systems often rely on large amounts of personal health data, there are legitimate concerns about the potential disclosure and misuse of sensitive information. Therefore, strong data encryption, secure storage and transparent privacy policies are essential to build trust among users and ensure compliance with data protection regulations.

6. Conclusion

The above analysis illustrates how to use machine learning techniques to remove baseline spectacle, power line interference, and muscle contraction to reduce ECG signal noise, how to use algorithms to diagnose and classify diseases, and how to use machine learning to perform functions such as disease prediction. The application of machine learning to the electrocardiogram holds great promise and will advance medicine in the future. But at the same time, there are still many shortcomings in terms of technology, efficiency and ethics for the realization of this technology. I hope that future papers can focus on these deficiencies to make up for. But overall, the feasibility of using machine learning technology in the analysis of ECG is still very high, and it is believed that this can be put into clinical medicine in the near future.

References

[1] F. Censi *et al*., "Effect of ECG filtering on time domain analysis of the P-wave," *2008 Computers in Cardiology*, Bologna, Italy, 2008, pp. 1077-1080, doi: 10.1109/CIC.2008.4749232.

[2] Mneimneh, M., Yaz, E., Johnson, M., & Povinelli, R. J. (2006). An adaptive kalman filter for removing baseline wandering in ECG signals. *Computing in Cardiology Conference*, 253–256. http://cinc.mit.edu/archives/2006/pdf/0253.pdf

[3] Mahesh S Chavan, R A Agrawala, M.D. Uplane. " Design and implementation of Digital FIR Equiripple Notch Filter on ECG Signal for removal of Power line Interference" 4th Wseas International Conference on Electronics, Control & Signal Processing Miami Florida USA 17-19 Nov. 2005 (pp 58-63).

[4] Sharma, N., & Sidhu, J. S. (2016). Removal of noise from ecg signal using adaptive filtering. *Indian Journal of Science and Technology*, *9*(48).

[5] Ramadevi, G. N., Rani, K. U., & Lavanya, D. (2015). Importance of feature extraction for classification of breast cancer datasets—a study. *International Journal of Scientific and Innovative Mathematical Research*, *3*(2), 763-368.

[6] Gosling, R., & Budge, M. (2003). Terminology for describing the elastic behavior of arteries. *Hypertension*, *41*(6), 1180–1182. https://doi.org/10.1161/01.hyp.0000072271.36866.2a

[7] Giuliani, A. (2017). The application of principal component analysis to drug discovery and biomedical data. *Drug Discovery Today*, *22*(7), 1069–1076. https://doi.org/10.1016/j.drudis.2017.01.005