The emergence and clinical significance of artificial intelligence–enhanced electrocardiography

Yong-Soo Baek

Division of Cardiology, Department of Internal Medicine, Inha University Hospital, Inha University College of Medicine, Incheon, Korea

The integration of artificial intelligence (AI) with electrocardiography (ECG), a technology known as AI-ECG, represents a transformative leap in the field of cardiovascular medicine. This innovative approach has significantly advanced the capabilities of ECG, traditionally used for diagnosing heart diseases. AI-ECG excels in detecting subtle changes and interconnected patterns in cardiac waveforms, offering a level of precision and sensitivity that was previously unattainable with conventional methods. The scope of AI-ECG extends beyond the realm of heart diseases. It has shown remarkable potential in predicting and identifying the impacts of noncardiac conditions on heart health, thereby broadening the diagnostic capabilities of ECG. This is especially valuable given the complex nature of cardiovascular diseases and their interactions with other health conditions. Despite its groundbreaking potential, AI-ECG faces several challenges. One of the primary concerns is the “black box” nature of AI algorithms, which can make the decision-making process opaque and difficult to interpret. This poses a challenge in medical settings where understanding the rationale behind a diagnosis is crucial. Additionally, the effectiveness of AI-ECG is dependent on the quality and diversity of the datasets used to train the algorithms. Limited or biased datasets can lead to inaccuracies and diminish the reliability of the technology. However, the benefits of AI-ECG are significant. It enables faster, more accurate diagnoses and has the potential to greatly enhance the efficiency of cardiovascular care. As research and technology continue to evolve, AI-ECG is poised to become an indispensable tool in the diagnosis and management of heart diseases.

Keywords: Artificial intelligence; Electrocardiography; Risk prediction; Prevention

INTRODUCTION

The year 2024 marks the centenary of the Nobel Prize in Physiology or Medicine being awarded to Willem Einthoven, a Dutch physician and physiologist who developed the first practical electrocardiography (ECG) method. Einthoven’s pioneering work with ECG, which introduced the PQRS (provocation, quality, region [or radiation], severity [or scale], and timing) terminology for cardiac waveforms, remains a cornerstone in the interpretation of electrocardiograms to this day. Over the years, the ECG has evolved, and in 1954, the American Heart Association officially introduced 12-lead ECG, which has become a widely used and easily accessible diagnostic tool for assessing cardiovascular conditions [1].

For over a century, ECG has been a fundamental tool in cardiovascular medicine, as it records the collective electrical activity of millions of cardiac muscle cells. The ECG pro-
procedure entails placing electrodes on the body’s surface to record the heart’s electrical activity, which is then depicted as waveforms. These waveforms are used to diagnose and manage heart diseases.

Traditionally, specific changes on an ECG, such as ST-segment elevation indicating myocardial infarction, T-wave alterations suggesting potassium abnormalities, and other noticeable ST-segment variations, have been used to identify particular clinical conditions. However, the sensitivity to detect subtle and interconnected changes present in ECG data was limited, as clinically significant changes needed to be substantial in size. Additionally, ECG interpretation requires considerable time and expertise, necessitating prolonged training.

Recent advancements in artificial intelligence (AI) and deep learning neural networks have revolutionized ECG analysis by enabling the discrimination of subtle differences in ECG signal waveforms. These developments have made it possible to detect nuanced and interconnected changes that were previously challenging to identify. As deep learning neural networks have become more sophisticated, research has expanded, demonstrating remarkable accuracy in various applications. These include identifying a person’s gender, detecting left ventricular dysfunction, discovering arrhythmias that are difficult for the human eye to discern in records, and even recognizing subtle ECG changes resulting from non-cardiac conditions that affect the heart’s electrical activity. The role of AI-enhanced ECG (AI-ECG) is evolving beyond traditional diagnostic functions. AI-ECG is emerging as a functional biological biomarker with a wide range of applications. This article aims to review and discuss the major research, achievements, and future prospects of AI in the context of 12-lead ECG.

THE HISTORY AND DEVELOPMENT OF AI

In 1950, British mathematician and computer scientist Alan Turing [2] introduced the Turing Test as a benchmark for evaluating machine intelligence, based on the ability of machines to engage in conversation with humans. The concept of AI has a surprisingly long history, with the term itself first coined by Professor John McCarthy of Dartmouth College in the United States during the 1956 Dartmouth Conference [3]. Despite early optimism, the field faced setbacks. Research into neural networks, which required complex calculations, was hindered by the limited capabilities of multilayer neural networks and the slow processing speeds of computers at the time, leading to two significant periods of stagnation.

In the late 2000s, AI research underwent a revival with the introduction of deep learning algorithms. These algorithms utilize multilayer neural networks to emulate the human brain’s functionality and tackle intricate problems. Notable deep learning algorithms include recurrent neural networks (RNNs), which are sequence models that process inputs and outputs in temporal sequences, and convolutional neural networks (CNNs), known for their exceptional performance in image processing. The advancement of deep learning algorithms, such as Google’s deep neural network (DNN) for image recognition, Facebook’s DeepFace, and Google DeepMind’s AlphaGo, in conjunction with the enhanced computing power provided by graphics processing units (GPUs) capable of parallel data processing, has fueled the current surge in AI. In recent years, the AI field has seen a shift towards generative models, exemplified by the generative pre-trained transformer (GPT) model.

In the realm of medicine, AI research in ECG analysis has been propelled by the availability of large ECG datasets and advanced computing hardware. Researchers have devised deep learning algorithms capable of analyzing ECG data with remarkable accuracy, which facilitates quicker diagnoses and heightens precision in detecting heart conditions. AI systems are also employed in predicting heart disease risk and in the screening, monitoring, and observation of patients. AI can be applied to ECG analysis in two primary ways: by automating tasks that are currently performed manually, such as identifying arrhythmias or acute myocardial infarction, and by extracting insights that surpass human perception, thus recognizing more complex clinical conditions.

The efficacy of AI software is typically evaluated using several performance metrics. These include accuracy, the confusion matrix, precision and recall, the receiver operating characteristic (ROC) curve, and the area under the ROC curve (AUC). Additionally, the F1 score, which is the harmonic mean of precision and recall, is frequently utilized in performance assessments.
ARRHYTHMIAS

The diagnosis and prediction of arrhythmias are among the most significant areas where the utility and value of AI-ECG are evident. Atrial fibrillation is the most common arrhythmia encountered in clinical practice [4]. In Korea, the prevalence of atrial fibrillation increased to 1.53% in 2015, with projections suggesting that 5.6% of the population may be affected by 2060 [5]. Atrial fibrillation is associated with an increased risk of death, heart failure, and is linked to approximately 20% to 30% of ischemic strokes [6]. It is a socially and economically burdensome cardiac disease, leading to cognitive impairment, decreased quality of life, and depression. Annually, 10% to 40% of patients with atrial fibrillation are hospitalized [7]. According to the recent EAST-AFNET 4 (Early Treatment of Atrial Fibrillation for Stroke Prevention Trial 4) and its subanalysis, published in the New England Journal of Medicine, early rhythm control in patients diagnosed with atrial fibrillation within 1 year, regardless of symptoms, significantly reduced mortality, hospitalization rates, and fatal complications by 21% [8]. Therefore, considerable efforts have been directed toward the diagnosis of atrial fibrillation, the most common arrhythmia, in the field of AI-ECG. Research on the diagnosis of other arrhythmias, such as ventricular tachycardia, has been relatively limited to date.

Automated electrocardiogram interpretation software that employs CNNs and RNNs within the realm of deep learning has been reported to accurately detect arrhythmias, including ongoing atrial fibrillation. Identifying the R peak is essential in AI analysis. Since atrial fibrillation, characterized by indistinct P waves, jitteriness, or irregular R-R intervals, can be relatively easily identified by the human eye, it is anticipated that AI can also accurately detect current instances of atrial fibrillation to some extent.

Attia et al. [9] presented a study in 2019 that trained a CNN AI method using 12-lead ECGs of normal sinus rhythm collected at the Mayo Clinic for 10-second durations. The AI-ECG algorithm’s performance was assessed on a self-test dataset, and it successfully identified patients with atrial fibrillation using normal sinus rhythm ECGs with an accuracy of 79%. Notably, the accuracy increased to 83% when using normal sinus rhythm ECGs from 1 month prior to the diagnosis of atrial fibrillation. This indicates that the electrical and structural remodeling associated with atrial fibrillation may begin before the condition is clinically diagnosed. Baek et al. [10] presented a study that trained a 12-lead ECG with normal sinus rhythm using an RNN-based method at the Korean Society of Holter and Non-invasive Electrocardiology in 2019. This innovative deep learning approach, which employs AI to identify specific features of atrial fibrillation and enhances data accuracy through classification by arrhythmia specialists instead of mechanical data labeling, more than doubled the F1 score compared to the results from the Mayo Clinic. Recent prospective studies have demonstrated that the use of AI-ECG for detecting atrial fibrillation results in an approximately threefold increase in sensitivity over traditional methods among high-risk patient groups [11]. While the diagnosis of arrhythmias has been the subject of extensive research, its direct impact on patient management has not been as thoroughly explored. Therefore, domestic researchers are currently undertaking large-scale, multi-institutional studies to investigate the relationship between AI-ECG and clinical outcomes beyond the mere diagnosis of atrial fibrillation [12].

Recent guidelines from the European Society of Cardiology [7] and the Korean Heart Rhythm Society [5] incorporate research on the use of normal sinus rhythm electrocardiograms, assisted by AI, to identify patients with paroxysmal atrial fibrillation. This is considered a significant advancement in the diagnosis of atrial fibrillation. Such research introduces a novel approach that could prove to be an invaluable tool for detecting atrial fibrillation, especially given the growing shortage of essential medical personnel and the critical importance of early detection. Hannun et al. [13] utilized a substantial dataset of electrocardiogram recordings to distinguish between 12 arrhythmia rhythm classes. Their findings demonstrated that the deep learning neural network outperformed cardiac internal medicine specialists in discernment, achieving an F1 score of 0.837. In another extensive deep learning study that analyzed Brazilian electrocardiogram big data, the researchers achieved an F1 score above 0.8 and specificity over 99% in classifying six types of arrhythmias, including atrial fibrillation. Furthermore, the detection of atrial fibrillation was strongly linked to clinical outcomes, serving as a powerful predictor of cardiovascular and all-cause mortality, with an increased risk noted in women [14].
Meanwhile, in research aimed at detecting ventricular tachyarrhythmias, which include ventricular tachycardia and ventricular fibrillation, the application of a DNN in automated external defibrillators has yielded favorable results. The DNN achieved an accuracy of 99.2%, a sensitivity of 98.8%, and a specificity of 99.3%, outperforming standard classification methods [15]. Recent studies using DNNs to analyze ECG signals have produced intriguing findings. These studies suggest the networks’ capability to detect patients at risk for ventricular tachyarrhythmias due to P-wave abnormalities, particularly for identifying life-threatening ventricular tachyarrhythmias in patients with hypertrophic cardiomyopathy [15]. Torsades de Pointes, a lethal ventricular arrhythmia, is linked to congenital or drug-induced long QT syndrome (LQTS). Recent research reported that a CNN model that quantified changes in ECG over a 10-second interval outperformed risk prediction based on the commonly used corrected QT interval (QTc), especially in distinguishing LQTS [16]. Furthermore, another study that developed a DNN using 12-lead ECG data found that the AI-ECG not only differentiated LQTS from QTc-based risk assessments but also accurately predicted the genetic subtypes of LQTS—types 1, 2, and 3—with nearly 80% accuracy [17]. As these studies continue to emerge and their accuracy is validated, they are expected to become valuable tools for predicting sudden cardiac events in the general population.

**STRUCTURAL HEART DISEASE**

AI-ECG has demonstrated considerable progress in the diagnosis of structural heart diseases. Left ventricular hypertrophy (LVH) serves as a marker for asymptomatic organ damage and is associated with an increased risk of cardiovascular diseases. Echocardiography remains the gold standard for diagnosing LVH. The assessment criteria for LVH based on ECG have traditionally been established by the voltage threshold of the RS peak.

A study conducted by researchers in Taiwan employed a back propagation neural network method and demonstrated high accuracy, precision, sensitivity, and specificity (0.96–0.97) in distinguishing LVH using 12-lead ECG [18]. The ECG shows promise as a screening tool, given its cost-effectiveness and rapid applicability in clinical settings. The Mayo Clinic developed a method to identify patients with asymptomatic left ventricular dysfunction by training a convolutional neural network with datasets from 44,959 patients using 12-lead ECG. The resulting AI-ECG exhibited high accuracy, sensitivity, and specificity in detecting heart failure in an independent cohort of 52,870 patients. Moreover, when AI screening indicated a positive result in patients without systolic dysfunction, their risk of developing future systolic dysfunction increased fourfold. These findings underscore the cost-effective and potent potential of AI-ECG as a screening tool for heart failure [19]. In Korea, researchers created an interpretable AI algorithm to detect heart failure with reduced ejection fraction (HFrEF) and validated its performance. The AI algorithm effectively detected HFrEF using both 12- and single-lead ECGs [20]. A recent meta-analysis of AI-ECG for heart failure detection corroborated the capability of ECG-based AI models to predict heart failure and left ventricular systolic dysfunction [21].

There has recently been an increase in the AI-ECG analysis for detecting structural heart diseases. Remarkable results have been reported for conditions such as pulmonary artery hypertension, severe aortic stenosis, severe mitral regurgitation, and cardiac amyloidosis [22–24]. One study demonstrated the potential of a DNN when applied to digital 12-lead ECGs from 2,448 patients with hypertrophic cardiomyopathy and a control group of 51,153 individuals matched for age and sex. The results were particularly promising for detecting hypertrophic cardiomyopathy in younger patients, with a sensitivity of 95% and a specificity of 92% [25]. Another study employed a technique known as long short-term memory, a form of RNN, to detect ST-segment elevation myocardial infarction. It achieved an accuracy of 0.987, an AUC of 0.997, and precision, recall, and F1 scores of 0.952, 0.870, and 0.909, respectively. Additionally, studies using a CNN model for detecting coronary artery disease showed excellent performance with an AUC of 0.869 [26,27]. The use of deep learning for cardiac diagnosis via electrocardiogram is highly promising. Future research may well concentrate on the early detection of rare cardiac conditions that are currently less understood clinically.

**NONCARDIAC DISEASES**

AI-ECG analysis has the potential to detect or monitor a range of diseases beyond those of the cardiovascular system. Its applications have been explored in the diagnosis of...
metabolic disorders, the detection of electrolyte imbalances, and the screening for conditions such as hyperthyroidism, anemia, and liver cirrhosis. Recent studies have yielded intriguing results, including the ability of AI-ECG analysis to quickly exclude SARS-CoV-2 infection with a negative predictive value of 99.2%. Furthermore, the use of AI-ECG to predict the severity of COVID-19 could help in the efficient allocation of medical resources during a pandemic [28]. Lin et al. [29] achieved high accuracy in distinguishing between hypokalemia (sensitivity, 96.7%; specificity, 93.3%; AUC, 0.926) and hyperkalemia (sensitivity, 83.3%; specificity, 97.8%; AUC, 0.958) using a deep learning model trained on a database of 66,321 ECGs. This model outperformed emergency room physicians and cardiologists in sensitivity, specificity, and accuracy on the test set.

AI-ECG utilizing 12-lead ECG has demonstrated the ability to predict a patient’s biological age. Intriguingly, it can also identify gender and predict 1-year mortality rates [30]. A recent study by Korean researchers, who analyzed 420,000 instances of 12-lead ECG data, showed that if the biological age predicted by AI-ECG exceeded the actual age by more than 6 years, there was a marked increase in the risk of major cardiovascular events, mortality, and hospitalizations related to cardiovascular issues [31].

LIMITATIONS

Several limitations of AI in this field should be kept in mind. Firstly, AI algorithms are often referred to as "black boxes" because of the opacity in their decision-making processes. The challenge of understanding the reasoning behind an algorithm’s conclusions is a significant obstacle to the broader adoption of AI systems, particularly in the medical field. Healthcare professionals must grasp the logic behind AI recommendations to build trust and promote patient engagement. Recently, efforts have been made to improve the interpretability of classification methods for ECG images. Techniques such as class activation map (CAM), Grad-CAM, SHAP (Shapley Additive Explanations), and the use of global weights importance in DNNs, along with other explainable AI (xAI) strategies, have been explored [32]. For instance, sensitivity maps in object recognition within ECG images can highlight the areas most associated with key electrocardiographic features in the classification decision, aligning closely with cardiologists’ diagnostic processes. However, interpreting deep learning models and justifying AI decisions is still a complex task due to AI’s inherent complexity. Secondly, most studies to date have been retrospective and conducted on limited datasets. There is a pressing need for large-scale prospective research, as well as validation and certification across various medical settings. Addressing issues such as imbalanced datasets and small patient cohorts is critical. Thirdly, despite the impressive performance of deep learning algorithms, the challenge of identifying optimal treatments and predicting outcomes while minimizing false positives and negatives is substantial. It is vital to be vigilant against overfitting, particularly considering the distinct characteristics of various diseases. Therefore, research that includes long-term follow-up may be required. Fourthly, the importance of accurate dataset labeling by experts during the design and development of AI algorithms cannot be overstated. Contaminated data can lead to performance degradation and introduce bias into AI systems.

CONCLUSIONS

AI-ECG has advanced to the point where it can detect subtle abnormalities in conditions that might otherwise be deemed "normal" by a cardiologist or a conventional ECG machine. Considering the critical importance of early diagnosis and management in heart disease, these AI techniques are particularly well-suited for the large-scale screening of ECGs and are anticipated to become a staple in clinical practice. The benefits of AI-enhanced ECGs include improved risk prediction, the ability to integrate with existing clinical variables, personalized treatment planning, and enhanced cost-effectiveness through the real-time analysis of ECGs. These advantages position AI as a potential ‘game changer’ in the diagnosis and management of cardiovascular disease in the near future.

ARTICLE INFORMATION

Conflicts of interest
The author has no conflicts of interest to declare.

Funding
The author received no financial support for this study.
REFERENCES


