Deep learning-based artificial intelligence for predicting risk and prognosis in patients with cardiovascular disease

Abstract:
Cardiovascular disease (CVD) is a major healthcare problem worldwide. Risk stratification and prognosis prediction are critical in identifying high-risk patients and in decision making to devise treatment strategies for patients with CVD. For this purpose, various models have been developed and validated against large amounts of population registry data by using conventional statistical methods such as regression-based models. However, these conventional models have a problem of over-generalization and are not applicable to all individual patients. Deep learning is a branch of artificial intelligence in which artificial neural networks are used to analyze data patterns; it is similar to functioning of the human neural system. An advantage of deep learning is the automatic learning of features and relationships from given data. Recently, deep learning achieved high performance in several medical domains, such as image classification, diagnosis, clinical outcome prediction, and gene analysis. The focus of this review is to summarize deep learning-based prediction models in patients with CVD in terms of accuracy in comparison with conventional models.

Keywords: Cardiovascular disease • Deep learning • Cardiac arrest • Acute myocardial infarction

Introduction
Cardiovascular Disease (CVD) is common in the general population and is a major healthcare problem worldwide [1]. Risk stratification and prognosis prediction to identify high-risk patients are essential in decision making to devise treatment strategies for patients with CVD. For this purpose, various risk-prediction models have been developed and validated against large amounts of population registry data by using conventional statistical methods such as regression-based models [2-4]. However, these conventional models use a limited number of pre-specified factors and have a problem of over-generalization, so call ‘one size does not fit all’.

Deep learning is a subset of machine learning in which artificial neural networks are used to analyze different factors; it has a structure similar to that of the human neural system [5]. Recently, there have been considerable advancements in deep learning–based Artificial Intelligence (AI), enabling the use of qualified big data and enhanced computing power. In the medical field, deep learning algorithms have achieved state-of-the-art performance in, for example, the discrimination of medical images and diagnosis or prediction models, by overcoming the limitations of conventional statistical methods [6-9].
In this review, we summarize recent achievements in deep learning–based AI for risk stratification and prognosis prediction in patients with CVDs.

**Risk Estimation and Prognosis Prediction Models in Cardiovascular Diseases**

Various risk-prediction models have been developed for estimating the risk of an initial CVD event in individuals without a documented CVD and for predicting the prognosis of patients diagnosed with CVDs. These conventional prediction models are based on regression models such as the logistic model [10], Cox hazard model [2], and Weibull model [3]. Models were developed based on a large pool of representative datasets and several variables (usually established risk factors) to predict the probability of cardiovascular events, such as cardiac death, myocardial infarction, and stroke, in terms of the odds ratio, relative risks, or hazard ratio [11].

The most widely known conventional risk-prediction model is probably the Framingham risk score [2]. It was proposed in 1998 and was derived from a largely Caucasian population in Europe, including 2,489 men and 2,856 women, in the age range of 30–74 years at the time of the Framingham Heart Study examination from 1971 to 1974. In this model, the probability of cardiovascular events is calculated using statistical tests including age-adjusted linear regression, logistic regression to test for trends, and age-adjusted Cox proportional hazards regression as well as its accompanying C-statistic. The model for risk estimation of stroke in Atrial Fibrillation (AF), the CHA2DS2-VASc score, is widely used for decision making regarding the use of anticoagulants in patients with AF [12]. Other well-established and validated risk-prediction models include the Systematic Coronary Risk Evaluation (SCORE) CVD death risk score for the 10-year risk of a first fatal atherosclerotic event [3], American College of Cardiology/American Heart Association (ACC/AHA) Atherosclerotic Cardiovascular Disease (ASCVD) risk estimator for the 10-year risk of heart disease and stroke [13], and Thrombolysis In Myocardial Infarction (TIMI) risk score [14] or Global Registry of Acute Coronary Events (GRACE) score for mortality prediction in ST Segment Elevation Myocardial Infarction (STEMI) patients [15].

Although these conventional models based on regression were useful in clinical practice, these statistical methods use a limited number of predictive factors that operate in the same manner for all patients. In particular, these models assume constant effects of risk factors for different ages and levels of other risk factors. Therefore, these models have a problem of over-generalization and are not applicable to all individual patients.

**Artificial Intelligence, Machine Learning, and Deep Learning**

AI is a term used to describe the application of computer science to simulate intelligent behavior and critical thinking for decision making in a manner comparable to a human being. The role of AI as a major catalyst in the healthcare revolution is unquestionable. AI has already started revolutionizing healthcare by leveraging big-data analysis to optimize healthcare services. It has made great advances in the field of disease diagnosis, drug development, personalized treatment, and improved gene editing [16–19].

Machine learning is an application of AI that provides computers the ability to automatically learn and improve from experience without explicit programming, human intervention, or assistance [20]. Deep learning is a type of machine learning that is inspired by the manner in which the human brain analyzes data. It combines computer science, statistics, and mathematical algorithms to find patterns and make decisions based on complex and big data. It includes feature learning, which is a set of methods that allows a model to be fed with raw data and to automatically identify the features and relationships needed for conducting a task [21,22].

It has been the most popular method for developing AI in recent years and has been empowered by big data and enhancements in computing power since 2010. In deep learning models, data are filtered through a cascade of multiple layers, with each successive layer using the output from the previous one to obtain its results [23,24].

**Deep Learning–Based AI Algorithm for CVD Prediction**

Recently, deep learning–based AI has achieved high performance in several medical domains, such as the detection of retinal disease, diagnosis of medical images, and electrocardiographic diagnosis of heart disease [25–28]. Predicting the risk and prognosis of CVD is a complex task with many factors to consider and needs time-consuming human operations to analyze that. Deep learning–based AI has excellent ability to solve problems automatically by analyzing these complex factors.

**CVD risk-prediction model**

Prediction models to estimate the risk of developing CVD require large amounts of complex data that are suitable for the application of deep learning. Cho et al. [29] developed a CVD prediction model using large-scale cohort data including 412,030 Korean adults in the National Health Insurance Service-Health Screening Cohort (NHIS-HEALS) for internal validation, 178,875 adults in the National Health Insurance Service–National Sample Cohort (NHIS-NSC) for the first external validation, and the 4,296 European adults in the Rotterdam Study [30] for the second external validation. In the external validation based on the Rotterdam Study, which included participants of different...
Review Article

169

ethnicities, the model demonstrated a C-statistic of 0.860 (0.824–0.897) in men and 0.867 (0.830–0.903) in women, as well as improved reclassification compared with conventional Cox regression (net reclassification index [NRI] of 36.9% in men and 31.8% in women).

Prediction model for mortality and prognosis of cardiac arrest

Cardiac arrest is a catastrophic event that leads to sudden cardiac death. It affects not only patients with CVD but also healthy people. Among patients with Return Of Spontaneous Circulation (ROSC), the in-hospital mortality rate is 50%-70%, and the majority sustain ischemic neurological injury [31,32].

Prediction of in-hospital cardiac arrest

Various medical conditions can cause in-hospital cardiac arrest, and the survival discharge rate of these patients is less than 20% [33]. More than half of in-hospital cardiac arrest cases result from respiratory failure or hypovolemic shock, and 80% of the patients who experienced cardiac arrest showed signs of deterioration in the eight hours before the cardiac arrest [34,35]. For predicting cardiac arrest, several risk-score models based on vital signs, including blood pressure and heart rate, are used for patient safety [36]. Kwon et al. [37] developed a deep learning–based early warning system (DEWS) to predict in-hospital cardiac arrest using vital sign data, including systolic blood pressure, heart rate, respiratory rate, and body temperature, from 52,131 patients. The area under the receiver operating characteristic curve (AUROC) of 0.850 and area under the precision-recall curve (AUPRC) of 0.044 for DEWS were significantly higher than those of a modified early warning score (AUROC of 0.603 and AUPRC of 0.003), which is one of the most widely used conventional approaches. An interesting point in this study is that the risk score of cardiac arrest in DEWS increased from 24 h before the cardiac arrest, and a DEWS >50% was found in patients 14 h before the cardiac arrest.

Prediction of mortality and neurologic outcomes for out-of-hospital cardiac arrest

Out-of-hospital cardiac arrest (OHCA) is a leading cause of global mortality. Globally, the percentage of survival to discharge was 5%-10%, depending on the region [38]. Even with successful resuscitation, ischemic neurological damage is inevitable in many cases. The prediction of prognosis is important in decision making to devise treatment strategies for patients with OHCA. Kwon and Jeon et al. [39] developed and validated a deep learning–based out-of-hospital cardiac-arrest prognostic system (DCAPS) for predicting neurologic recovery and survival to discharge. The deep learning model was developed using data from the Korea OHCA Registry (KOHCAR) in South Korea, in which 36,190 patients with OHCA from 712 emergency departments (EDs) were enrolled (Figure 1) [39]. As predictor variables during model development, the authors utilized only the information available at the time of ROSC, including age, sex, place of OHCA, etiology of OHCA (disease or trauma), ROSC in emergency medical service (EMS), event witness, bystander cardiopulmonary resuscitation (CPR), initial electrocardiography (ECG) rhythm of EMS, initial ECG rhythm of ED, and time from ED visit to ROSC. The AUROC of DCAPS was 0.953 (95% confidence interval [CI], 0.952–0.954) for neurologic recovery, which is significantly higher than the value of 0.817 achieved using the conventional model (95% CI, 0.815–0.820). In terms of survival to discharge, the AUROC of the DCAPS and conventional model were 0.901 (95% CI, 0.900–0.903) and 0.736 (95% CI, 0.734-0.739) (Figure 2) [43]. Therefore, deep learning–based AI model accurately predicted the neurologic recovery and survival to discharge of OHCA patients, outperforming the conventional method.

Figure 1: Development and validation of deep learning–based prognostic model. Reprinted from Kwon et al., Resuscitation 139: 84-91 (2019), Copyright by Elsevier
Prediction of mortality in patients with AMI

Many efforts have been made to accurately predict the prognosis of patients with acute myocardial infarction (AMI). Conventional risk scoring systems, including TIMI [10,14], GRACE [15], and the acute coronary treatment and intervention outcomes network (ACTION) [40], are widely validated and accepted scores that are estimated using patients’ clinical information. However, these prognostic models have limitations in current daily practice. Firstly, these systems are questionable in contemporary practice because they had been developed 20 years ago. Additionally, as these models use only selective variables based on a conventional statistical method, there is a possibility of loss of important information. Shouval et al. [41] developed a machine-learning algorithm to predict 30-day mortality following STEMI and compared it with the conventional GRACE and TIMI scoring systems. The best accuracy achieved using the machine-learning algorithm (AUC of 0.91 and standard deviation [SD] of 0.04) was similar to that of GRACE (AUC of 0.87 and SD of 0.06) and better than that of TIMI (AUC of 0.82, SD of 0.06, and P<0.05). In a study on deep learning–based risk stratification for mortality of patients with AMI (DAMI) [42], a deep learning algorithm was developed from 22,875 AMI patients from the Korean working group of acute myocardial infarction (KorMI) registry. The algorithm used a total of 37 variables of demographic and laboratory data. During the accuracy test for STEMI patients, the AUC of DAMI was 0.905 (95% CI, 0.902–0.909), and this result significantly outperformed the GRACE score (0.851 with 95% CI, 0.846–0.856), ACTION score (0.852 with 95% CI 0.847–0.857), and TIMI score (0.781 with 95% CI 0.775–0.787). The algorithm also showed better accuracy than conventional models in a non-ST segment elevation myocardial infarction (NSTEMI) patient group. Figure 3 shows the results of reclassification of individuals predicted to be in the intermediate-risk group through additional DAMI assessment [39]. Among the 3,526 patients who were placed in the intermediate-risk group based on the GRACE score, 1937 patients were reclassified into the low-risk group. Furthermore, among 50 patients who met with in-hospital death, 24 patients were reclassified into the high-risk group based on the DAMI score. This implies that DAMI can differentiate the mortality risk of patients with AMI more sensitively than conventional models.
Limitation of Deep Learning

AI based on deep learning has enormous potential in the medical field, and it could improve the accuracy of diagnosis and support clinical decisions for many diseases. However, it is also necessary to clearly recognize the limitations of AI and make efforts to overcome them. One of the most important characteristics of deep learning is that it does not use any medical knowledge; rather, it uses only the relationship among variables of given data and can easily overfit the dataset. Therefore, external validation is essential in medical AI research. Here, the term “external” refers not only to exclusively separated dataset, but also to data from different environments. Validation of an established AI model in a completely new dataset is mandatory for AI research to overcome overfitting. For that, cross-validation is one of the preferred methods to reduce the variance in prediction error and to give an insight on how the model will generalize to an independent dataset [22]. The second limitation is that current AI technology cannot reveal the decision process of deep learning, which is so-called a black box. In other words, although we can develop deep learning–based AI by fitting each coefficient, we cannot interpret the AI in terms of its approach to decision making. Interpretable deep learning has been studied recently, but it takes much time to apply the theory to the medical field. As with past medical research, studies on AI must not only provide accurate results but also attempt to analyze and understand AI.

Conclusion

We are currently on the brink of the “fourth industrial revolution,” a technological revolution that will fundamentally change the patterns by which we live, work, and relate to one another. Deep learning has achieved state-of-the-art performance in several medical domains and outperforms existing conventional methods. Deep learning showed better performance than conventional models in risk stratification and prognosis prediction for CVD and could be of great help in the evaluation and treatment of patients with CVD in the future.

Conflict of Interest

The authors declare no conflict of interest.

Funding

This study received no funding.

Author’s contribution

Ki-Hyun Jeon and Joon-myoung Kwon contributed equally to this article.
Review Article

References


