Approaches of Automated Heart Disease Prediction

Subjects: Computer Science, Artificial Intelligence | Computer Science, Interdisciplinary Applications

Contributor: Cătălina-Lucia Cocianu , Cristian Răzvan Uscatu , Konstantinos Kofidis , Sorin Muraru , Alin Gabriel Văduva

Cardiovascular diseases (CVDs) are the leading cause of death globally. Detecting this kind of disease represents the principal concern of many scientists, and techniques belonging to various fields have been developed to attain accurate predictions.

Cardiovascular diseases deep learning

1. Introduction

Cardiovascular diseases (CVDs) are one of the main causes of the rising mortality rate all around the world, with an unhealthy diet, alcohol consumption, smoking, and a lack of physical activity contributing to the risk of developing such conditions. CVDs are a class of disorders of the heart and blood vessels mainly consisting of coronary heart, cerebrovascular, and rheumatic heart disease, respectively. CVDs take an estimated 17.9 million lives each year, which is an estimated 32% of all deaths worldwide [1]. As symptoms can often be similar to those of other illnesses and age-related issues, it is difficult for medical professionals to diagnose them. Finding the people at highest risk of CVDs, early diagnosis and appropriately treating those suffering from this kind of disease can prevent premature deaths.

Detecting CVDs represents the principal concern of many scientists in the artificial intelligence area, and various techniques have been developed for attaining a method that can perform this accurately, such as classical statistical strategies, machine learning, deep learning-oriented algorithms, and evolutionary computation-based mechanisms together with various data mining pre-processing methods.

2. Classical ML Techniques

In [2], three different approaches based on classical statistical methods were used to identify CVDs. The first approach is based on the usage of various machine learning (ML) algorithms, such as random forest, logistic regression, K-Nearest Neighbors (KNN), support vector machine (SVM), decision tree, and XGBoost, using the UCI Heart Disease dataset. Note that feature selection and outliers detection were not considered. The second approach used only the feature selection mechanism, while the third approach brought both feature selection and outliers detection into practice. The most accurate algorithm was obtained in the last approach by KNN, with the percentage of correct classification being 84.86%.

The research reported in [3] also uses statistical models such as SVM, Gaussian Naïve Bayes (GNB), logistic regression, LightGBM, CGBoost, and random forest (RF) to create a classifier that is tested and trained on the Cleveland Clinic Foundation for Heart Disease dataset. The accuracy of the models is measured using performance matrices and confusion matrices. The experimentally established results pointed out that the random forest classifier achieved the best accuracy, followed by SVM and logistic regression.

Various ML techniques, including SVMs, decision trees (DTs), and Naïve Bayes (NB), were used in [4] to generate diagnosis results on the South African Heart Disease dataset. The performance of these models was compared according to accuracy, sensitivity, specificity, true positives (TPs), true negatives (TNs), false positives (FPs), and false negatives (FNs). Results indicated that NB had the highest accuracy rate; however, it was not satisfactory in terms of specificity and sensitivity. On the contrary, SVMs and DTs provided higher specificity ratings but displayed inadequate sensitivity. Thus, it was concluded that further research is necessary to elevate the performance and increase sensitivity and specificity scores.

Additionally, the research work reported in [5] has demonstrated the efficacy of using a cost-sensitive ensemble method for the diagnosis of heart diseases. To assess the performance of this approach, the Statlog, Cleveland, and Hungarian heart disease datasets were selected for analysis. Furthermore, various metrics, such as E, MC, G-mean, precision, recall, specificity, and AUC, were used to measure the effectiveness of the classifiers. Relief algorithms were employed to identify the most pertinent features and eliminate any effects of irrelevant features. This study has provided promising results and is a step forward in the development of more sophisticated classifiers with improved accuracy when used in combination with new algorithms.

Recent research has demonstrated the utility of ML techniques in forecasting the 90-day prognosis of individuals diagnosed with transient ischemic attack and minor stroke. The study, conducted in [6], utilized data from the CNSR-III prospective registry study, which included demographic, physiological, and medical history information of patients with the medical condition. The authors found that models constructed using logistic regression and machine learning exhibited superior performance, as evidenced by their Area Under the Curve (AUC) measure exceeding 0.8. Of the models employed, the Catboost model demonstrated the highest AUC score at 0.839.

3. Deep Learning Techniques

With the development of computing power appeared the chance to use more resource demanding algorithms, which are a subcategory of machine learning models named deep learning (DL). Those algorithms have evolved rapidly in recent years and have proven useful, with robust results for various projects, in different areas of interest.

Diagnosing diseases is a good example of showcasing the abilities of deep learning models and the different areas they excel in classification problems. In [7], the possibility of categorizing and understanding MRI scans to diagnose brain tumors is studied. By comparing two models, the convolutional neural networks (CNN) and a deep neural network (DNN), which is typically a feed-forward network and makes it a perfect fit for these types of problems, it is concluded that the latter provides the most accurate results.

The authors of [8] proposed the usage of multiple data mining and deep learning techniques, using a dataset containing information selected by taking into account the history of patients' heart problems in correlation with other medical aspects. Experimentally established results showed that the best classification scores of the proposed ML approach were obtained by the random forest classifier: accuracy (90.21%), precision (90.22%), recall (90.21%), and F1 score (90.21%).

In [9], a CNN-based diagnosis system was proposed. This model was found to be comparably effective to traditional machine learning models, such as SVM and random forests, particularly in predicting negative cases, meaning those cases without coronary heart disease. The architecture of the proposed model was a sequential feedforward one-input–one-output network, which begins with the application of LASSO regression, which adds a penalty and eliminates coefficients that help control true negatives in the dataset.

There has been research that used both machine and deep learning algorithmic approaches to classify and create predictions for heart diseases. It has been proven that taking into account the medical history of the patients allows deep learning models to outperform machine learning algorithms and yield high accuracy [8].

4. Hybrid Methods

A hybrid method aiming to extract significant features using ML techniques for CVDs prediction was reported in [10]. The classification model was developed with various feature combinations and is based on the aggregation of random forest and linear classification models. The proposed algorithm, hybrid random forest with a linear model (HRFLM), was assessed to have an 88.7% accuracy. The study also points to the idea that new feature selection methods and new combinations of ML techniques can be used to achieve highly accurate classification algorithms.

Another method to classify patients suffering from CVDs is reported in [11]. The method is based on the usage of ML techniques and ontology to build an efficient model capable of accurately predicting the presence of cardiac disease and facilitating early diagnosis. The main purpose is to extract the relevant rules from the DT algorithm, then to implement these rules in an ontology using the Semantic Web Rule Language (SWRL). The model reached a level of accuracy of 75% and an F1 score of 80.5%, outperforming the standard DT model (73.1% accuracy, 73.8% F1 score).

In the case study presented in [12], the usage of evolutionary algorithms (EAs), such as genetic algorithms (GAs) and particle swarm optimization (PSO), was proved to raise the overall accuracy of ML algorithms. The reported research combines EAs with Naïve Bayes and support vector machine for feature selection. The most successful algorithm in terms of classification accuracy uses GA as a feature extraction strategy.

Various techniques can be used to enhance deep learning algorithms to increase their accuracy. An example is the combination of a Multilayer Perceptron (MLP) algorithm with the use of the Back Propagation of Errors Algorithm, which fine tunes the weights based on the error rate acquired from the previous iteration [13]. It was experimentally

proved that the new model has improved performance compared to similar approaches, thus creating a better tuned MLP model for the use of classification problems.

Another way to optimize the algorithms is to use swarm intelligence optimization techniques. Such a technique is particle swarm optimization (PSO) which helps the model to determine the optimal weight and bias values. Combining this technique with a MLP model and testing on the heart disease dataset gives a model that outperformed the initial model [14].

In [15], various techniques for recognizing cardiovascular diseases (CVDs) were examined, including Data Mining (DM), DL, ML, and Soft Computing. The authors reviewed the literature on CVD recognition and presented the findings in terms of advantages, limitations, and accuracy levels. The results revealed that certain approaches, such as the utilization of DM and GAs, yielded high accuracy scores of around 96%. However, other methods produced lower accuracy levels, with accuracy scores of around 45%. The best accuracy was observed in the use of Neural Networks (NN), which achieved a score of 99%.

The capabilities of different methods in predicting heart disease can be further enhanced by analyzing the data obtained from other sources such as electrocardiograms (ECGs). ECG waves are widely used to diagnose cardiovascular illnesses. The work reported in [16] aims to develop a non-linear vector decomposed neural network (NVDN) to classify ECG data. The proposed method was tested using well-known datasets from UCI and Physio net. After denoising the images with the use of frequency wavelet decay strategies, a subset of common features is identified. The NVDN model is then used to predict CVDs. The model produces decent results in terms of the F1 score, accuracy, sensitivity, and specificity; however, the forecasts are inaccurate. It is thus proposed to minimize time complexity and improve categorization.

In [17], neuro-fuzzy systems were used to learn predictive models from training data to create decision rules meant to support the decision-making process in cardiovascular risk assessment. The reported accuracy has reached 0.91, proving that artificial intelligence models are a valuable help for clinicians.

The literature review shows that there are many possibilities and opportunities for optimization techniques and different approaches to improve the accuracy of the deep learning models, which makes them even better candidates for classification problems.

Recent studies aiming to increase the accuracy of the diagnosis systems indicate that hyperparameter optimization techniques proved powerful tools. In [18], a radial basis function neural network (RBFNN) designed to identify and diagnose non-linear systems is proposed. The hyperparameters of the RBFNN are computed using PSO-based techniques. The resulting algorithm, which incorporates a spiral search mechanism, proves to improve prediction accuracy and can be extended to various types of neural networks. A PSO algorithm is also used for parameter optimization in [19]. The proposed diagnostic system is based on CNN and recognizes malignant and benign people in the attempt of identifying early-stage breast cancer.

A novel ensemble technique combining NB, DT, and SVM is introduced in [20] to classify heart diseases. The proposed approach involves two layers of base learners and a final meta-learner used to optimize the prediction accuracy. Other approaches involving a more accurate representation of neuronal activity use spiking neural networks [21]. To optimize the recognition rate, various heuristic algorithms including Cuckoo Search Algorithm, Grasshopper Optimization Algorithm, and Polar Bears Algorithm are used to compute the parameters of the spiking NN.

In the research, new methods that improve the accuracy of standard classifiers by combining them with hyperparameter optimization algorithms are proposed. The hybrid approaches are based on the LSTMs and SVMs models due to their promising classification performance and generalization capacity. The soft-margin nonlinear SVM model is used together with the basic evolutionary strategy to tune the hyperparameters. The second attempt is aimed at improving the accuracy of a DNN model that is based on a couple of LSTM layers. In this case, the aim is to optimize the classification accuracy by setting the activation function, the number of hidden neurons corresponding to each LSTM layer, and the dropout rate corresponding to the dropout layer. The computation is carried out by one of the most successful Bayesian optimization algorithms, namely tree-structured Parzen estimation (TPE). To establish meaningful conclusions, various tests were designed and the performance of each classifier was measured using the MSE indicator and the F1 score. The tests compared the most commonly used classifiers and the DNN models against the proposed hybrid algorithms. The Cleveland Clinic Foundation dataset together with its extension Statlog was used in the tests. In the case of the Cleveland Clinic Foundation set, the best results were obtained by the proposed 2MES-SVM, which improves SVM results by 2.5%. In the case of the aggregate dataset, the best results were obtained when using the proposed TPE-LSTM algorithm, with the F1 score being improved by 3.7%.

References

1. WHO. CVD Death Estimation. Available online: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) (accessed on 1 September 2022).
2. Bharti, R.; Khamparia, A.; Shabaz, M.; Dhiman, G.; Pande, S.; Singh, P. Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning. Comput. Intell. Neurosci. 2021, 2021, 8387680.
3. Karthick, K.; Aruna, S.K.; Samikannu, R.; Kuppusamy, R.; Teekaraman, Y.; Thelkar, A.R. Implementation of a Heart Disease Risk Prediction Model Using. Comput. Math. Methods Med. 2022, 2022, 6517716.
4. Gonsalves, A.H.; Fadi, T.; Rami Mustafa, M.A.; Singh, G. Prediction of Coronary Heart Disease using Machine Learning: An Experimental Analysis. In Proceedings of the 2019 3rd International Conference, Xiamen, China, 5–7 July 2019; pp. 51–56.

5. Qi, Z.; Zhang, Z. A hybrid cost-sensitive ensemble for heart disease prediction. BMC Med. Inform. Decis. Mak. 2021, 21, 73.
6. Chen, S.-D.; You, J.; Yang, X.-M.; Gu, H.-Q.; Huang, X.-Y.; Liu, H.; Feng, J.-F.; Jiang, Y.; Wang, Y.-J. Machine learning is an effective method to predict the 90-day prognosis of patients with transient ischemic attack and minor stroke. BMC Med. Res. Methodol. 2022, 22, 195.
7. Mohsen, H.; El-Dahshan, E.S.A.; El-Horbaty, E.S.M.; Salem, A.B.M. Classification using deep learning neural networks for brain tumors. Future Comput. Inform. J. 2018, 3, 68–71.
8. Barhoom, A.M.; Almasri, A.; Abu-Nasser, B.S.; Abu-Naser, S.S. Prediction of Heart Disease Using a Collection of Machine and Deep Learning Algorithms. Int. J. Eng. Inf. Syst. (IJEAIS) 2022, 6, 13.
9. Dutta, A.; Tamal, B.; Meheli, B.; Acton, S.T. An Efficient Convolutional Neural Network for Coronary Heart Disease Prediction. Expert Syst. Appl. 2020, 159, 113408.
10. Mohan, S.; Thirumalai, C.; Srivastava, G. Effective heart disease prediction using hybrid machine learning techniques. IEEE Access 2019, 7, 81542–81554.
11. El Massari, H.; Gherabi, N.; Mhammedi, S.; Sabouri, Z.; Ghandi, H. ONTOLOGY-BASED DECISION TREE MODEL FOR PREDICTION OF CARDIOVASCULAR DISEASE. Indian J. Comput. Sci. Eng. 2022, 13, 851–859.
12. Aleem, A.; Prateek, G.; Kumar, N. Improving Heart Disease Prediction Using Feature Selection Through Genetic Algorithm. Commun. Comput. Inf. Sci. 2022, 1534, 765–776.
13. Durairaj, M.; Revathi, V. Prediction Of Heart Disease Using Back Propagation MLP Algorithm. Int. J. Sci. Technol. Res. 2015, 4, 235–239.
14. Al Bataineh, A.; Manacek, S. MLP-PSO Hybrid Algorithm for Heart Disease Prediction. J. Pers. Med. 2022, 12, 1208.
15. Srivastava, K.; Choubey, D.K. Soft Computing, Data Mining, and Machine Learning Approaches in Detection. In Advances in Intelligent Systems and Computing, Proceedings of the 19th International Conference on Hybrid Intelligent Systems, Bhopal, India, 10–12 December 2019; Springer: Cham, Switzerland, 2019; pp. 165–175.
16. Suhail, M.M.; Razak, T.A. Cardiac disease detection from ECG signal using discrete wavelet transform with machine learning method. Diabetes Res. Clin. Pract. 2022, 187, 109852.
17. Casalino, G.; Castellano, G.; Kaymak, U.; Zaza, G. Balancing Accuracy and Interpretability through Neuro-Fuzzy Models for Cardiovascular Risk Assessment. In Proceedings of the 2021 IEEE Symposium Series on Computational Intelligence (SSCI), Orlando, FL, USA, 5–7 December 2021; pp. 1–8.
18. Ahmad, Z.; Li, J.; Mahmood, T. Adaptive Hyperparameter Fine-Tuning for Boosting the Robustness and Quality of the Particle Swarm Optimization Algorithm for Non-Linear RBF Neural

Network Modelling and Its Applications. Mathematics 2023, 11, 242.

19. Ogundokun, R.O.; Misra, S.; Douglas, M.; Damaševičius, R.; Maskeliūnas, R. Medical Internet-of-Things Based Breast Cancer Diagnosis Using Hyperparameter-Optimized Neural Networks. Future Internet 2022, 14, 153.
20. Prakash, V.J.; Karthikeyan, N.K. Dual-Layer Deep Ensemble Techniques for Classifying Heart Disease. Inf. Technol. Control. 2022, 51, 158–179.
21. Połap, D.; Woźniak, M.; Hołubowski, W.; Damaševičius, R. A heuristic approach to the hyperparameters in training spiking neural networks using spike-timing-dependent plasticity. Neural Comput. Appl. 2022, 34, 13187–13200.

Retrieved from https://encyclopedia.pub/entry/history/show/114887