

A Review of Feature Selection and Classification Approaches for Heart Disease Prediction

Fathania Firwan Firdaus¹, Hanung Adi Nugroho², Indah Soesanti³

Abstract—Cardiovascular disease has been the number one illness to cause death in the world for years. As information technology develops, many researchers have conducted studies on a computer-assisted diagnosis for heart disease. Predicting heart disease using a computer-assisted system can reduce time and costs. Feature selection can be used to choose the most relevant variables for heart disease. It includes filter, wrapper, embedded, and hybrid. The filter method excels in computation speed. The wrapper and embedded methods consider feature dependencies and interact with classifiers. The hybrid method takes advantage of several methods. Classification is a data mining technique to predict heart disease. It includes traditional machine learning, ensemble learning, hybrid, and deep learning. Traditional machine learning uses a specific algorithm. The ensemble learning combines the predictions of multiple classifiers to improve the performance of a single classifier. The hybrid approach combines some techniques and takes advantage of each method. Deep learning does not require a predetermined feature engineering. This research provides an overview of feature selection and classification methods for the prediction of heart disease in the last ten years. Thus, it can be used as a reference in choosing a method for heart disease prediction for future research.

Keyword—Feature Selection, Classification, Heart Disease.

I. INTRODUCTION

Cardiovascular disease is the leading cause of death in the world. It is a disease that is related to the condition of the heart and blood vessels. The risk factors for heart disease include raised blood pressure, raised blood glucose, raised blood fat, and obesity [1]. Early detection of heart disease can be helpful for patients to get early treatment. Along with the development of information technology, researchers have developed a computer-assisted diagnosis system for heart disease. Predicting heart disease using a computer-assisted system can help in reducing the patient's waiting times and costs [2].

In order to choose the most relevant variables for heart disease, a feature selection method can be used [3]. Feature selection removes irrelevant, redundant, and noisy data; thereby, it can reduce storage, computational costs and avoid a decrease in learning algorithm performance [4]. Besides removing excessive features, feature selection also aims to improve prediction accuracy and reduce analysis time [5]. Feature selection approaches include filter, wrapper, embedded, and hybrid.

Classification requires the use of machine learning to predict data instances included in certain group members [6]. The

classification uses training data to build the prediction model; then, the resulting model is used to predict the problem using the test data [7]. Several classification approaches have been applied to the heart disease dataset for predicting heart disease, including traditional machine learning, ensemble learning, hybrid, and deep learning.

The purpose of this study is to give an overview of the advantages and disadvantages of the feature selection method and to review the performance of several classification methods. Thus, beginners can study and choose a suitable method for the prediction of heart disease. This research is organized as follows. First, feature selection approaches for the prediction of heart disease are presented in Section II. Secondly, classification approaches for predicting heart disease are presented in Section III. Finally, Section IV concludes the topics that have been discussed.

II. FEATURE SELECTION APPROACHES FOR HEART DISEASE PREDICTION

Feature selection is a pre-processing technique in data mining to select relevant features and reduce data by removing unnecessary and redundant attributes from the dataset [7]. The purpose of feature selection is to get better prediction performance, accelerate the prediction process, reduce costs, and understand the process of generating data better [8]. The researchers have implemented various feature selection approaches to obtain optimal prediction results. There are several feature selection categories, namely filter, wrapper, embedded, and hybrid.

A. Filter

The filter method selects variables based on rank and removes variables that are below the threshold [9]. This method uses ranking technique criteria since it is simple and produces relevant features [10]. The advantage of the filter method is that it is fast, scalable to large datasets, and independent [11]. The drawback of the filter method includes ignoring feature dependencies and lacking interaction with the classifier [12]. It considers features separately, which leads to reduce the classification performance.

The information gain [13] selects features by sorting the weight of the feature using maximum entropy. In the experiment of heart disease classification using information gain feature selection and neural network, the classifier achieved training accuracy of 89.56% and validation accuracy of 80.99% [13]. Information gain can decrease the complexity of the model, increase computational efficiency, and improve prediction accuracy. However, information gain is not suitable to be applied to attributes with a large number of distinct values.

^{1,2,3} Department of Electrical and Information Engineering, Faculty of Engineering, Universitas Gadjah Mada, Yogyakarta 55281 Indonesia (e-mail: fathaniafirwan@mail.ugm.ac.id, adinugroho@ugm.ac.id, indahsoesanti@ugm.ac.id)

ReliefF [14] is a filter-based feature selection that chooses features by a high-quality rank score. On the Statlog heart dataset, using ReliefF and Support Vector Machine (SVM) to predict heart disease produced an accuracy of 84.81% [14]. It can shorten the processing time of some classifiers, such as K-Nearest Neighbor (K-NN), Multilayer Perceptron (MLP), and Support Vector Machine (SVM). ReliefF improves the basic Relief in solving noise problems more reliably. Furthermore, it is preferable on the voluminous database than the wrapper method since it is faster in execution.

Minimum Redundancy Maximum Relevance Feature Selection (MRMR) [15] produces feature selection in the form of weights. Features are ranked by selecting a subset of features based on their weightiness and similarity. The heart disease prediction experiment using MRMR and SVM achieved an accuracy of 84.85% [15]. MRMR solves the problem of feature set redundancy that occurs in a simple ranking approach because of the correlated features. MRMR can choose features that are the most relevant to the target class and are minimally redundant to themselves.

Cuckoo Search Algorithm (CSA) filter-based feature selection [16] adapts the general filter method and enhances it with the cuckoo-inspired algorithms. It is inspired by the age of the cuckoo birds. The heart disease prediction using CSA and SVM achieved an accuracy of 89.90% on the Eric dataset, 94.22% on the Hungarian dataset, 94% on the Statlog dataset, and 90% on the Z-Alizadeh Sani dataset [16]. While on the echocardiogram dataset, heart disease prediction using CSA and Naïve Bayes achieved the highest accuracy of 100% [16]. CSA filter-based can reduce the features and improve the heart disease prediction accuracy. However, there is no dependency between the selected features. As such, a Cuckoo Search Algorithm (CSA) based on wrapper can be developed to increase the dependability of features.

B. Wrapper

The wrapper method generates a subset of features using a search technique and evaluates the subset using a supervised learning algorithm in terms of misclassification and accuracy [17]. Wrapper resolves the problems in the filter method. It can interact with the classifier and think of dependencies between features [11]. The process of wrapper feature selection is optimized for the classification algorithm used. Therefore, generally, it results in better performance accuracy over the filter method. However, this method is computationally expensive and more complex than the filter method and tends to over-fitting in small training datasets [7].

Genetic Algorithm (GA) [18] explores the subset space set to get features that maximize prediction accuracy and minimize irrelevant attributes. Genetic Algorithm (GA) is inspired by the process of natural selection. There are five essential problems in the GA, namely chromosome encoding, selection mechanisms, fitness evaluation, genetic operators, and criteria to stop the GA. Using GA and Linear Discriminant Analysis (LDA) on the Cleveland heart disease dataset achieved an accuracy of 89.07% for binary class and 67.22% for multiclass [18]. Genetic Algorithm (GA) feature selection can improve classification accuracy both in binary class and multiclass for

the prediction of heart disease. A Genetic Algorithm can manage a dataset with many features, but it is computationally expensive since each individual evaluation requires model training.

Forward selection [19] begins with a blank set of features, then selects features by adding variables one by one until it meets certain criteria. In the heart disease prediction research using forward selection and K-NN classifier, the prediction accuracy improved up to 78.66% compared to the K-NN without forward selection, which yielded an accuracy of 73.44% [19]. Forward selection can select the most significant features and improve the prediction accuracy, but the features that have been selected in the previous step cannot be removed.

Sequential Backward Selection (SBS) [20] concludes the feature space to be a feature subspace with low classifier performance latency. SBS sequentially removes features based on the objective function of all features until there are enough features in the new feature subspace. An experiment using SBS feature selection and K-NN classifier produced an accuracy of 90% in predicting heart disease [20]. The drawback of SBS is that it cannot reconsider a feature that has been removed from the feature subset.

C. Embedded

The embedded method looks for a subset of features that are optimal for a specific classification algorithm when building a classifier [21]. The embedded method can interact with the classifier, pay attention to dependencies between features, and has better computational complexity than the wrapper [12]. The embedded method is computationally cost-effective and less over-fitting than the wrapper method; however, it is computationally costlier than the filter method. This method makes decisions depending on the classifier, so the choice of features is influenced by the classifier hypothesis and not suitable with some other classifiers [7].

Artificial Neural Network (ANN) [22] is a simulation of the human brain using experiential learning. Artificial Neural Network (ANN) feature selection performs feature removal several times until the model with the selected subset of features achieves optimal results. In the study, a multilayer perceptron neural network was used for selecting features from the Ischemic heart dataset, and a neural network was used to classify the heart disease, which yielded training accuracy of 89.4% and testing accuracy of 82.2% [22]. This method is able to get optimal features, which can help in optimizing certain learning algorithms. However, this method performs the selection of features in iterations, which is unsuitable for voluminous data since it can be computationally costlier.

The decision tree [23] allows adding or removing attributes to get better results and accuracy. The heart disease prediction model using the gain ratio in the decision tree feature selection and decision tree classifier achieved an accuracy of 85% [23]. This method chooses the separation attribute that minimizes the entropy value, thereby maximizing the information gain. The benefit of this method is it gives structural information of the features, so it can detect features that are important for classification. However, this method depends on a certain learning algorithm.

Support Vector Machine-Recursive Feature Elimination (SVM-RFE) [24] eliminates one feature at a time using sequential backward elimination. The first features to remove include the noise, redundancy, or irrelevant feature; the last feature to remove is the most relevant feature. The heart disease prediction using SVM-RFE and Principal Component Analysis-Support Vector Machine (PCA-SVM) yielded an accuracy of 88.24% [24]. This method can improve heart disease prediction accuracy by reducing redundant information. However, it is a greedy method that tries to find the best possible combination for classification by removing the worst features one by one. The features that are removed earlier when combined with other features might provide a significant performance improvement.

Variable Precision Rough Set (VPRS) [3] is an extension of the classical rough set. VPRS feature selection with the combination VPRS and Repeated Incremental Pruning Error Reduction (RIPPER) classifier in the Cleveland heart disease dataset yielded an accuracy of 88.89% [3]. However, VPRS feature selection with VPRS or RIPPER classifier individually results in lower accuracy. This method can find the optimal features for classification and interact with the classifier, but only suitable for certain classifiers.

D. Hybrid

The hybrid method combines several feature selection approaches and exploits the advantages of each approach to get optimal results. The hybrid method that is widely used is a combination of filter and wrapper. Combining filter and wrapper can help to improve filter accuracy and reduce wrapper processing time [25]. The hybrid method has a better performance accuracy than a filter, less computational complexity, and less over-fitting compared to a wrapper, but only matches certain classifiers [12].

Particle Swarm Optimization-Support Vector Machine (PSO-SVM) [26] uses the diversity function and tuning function to maintain algorithm consistency. In the experiment using the Cleveland heart disease dataset, the PSO-SVM feature selection with SVM classifiers has better performance than using traditional PSO feature selection, which yielded an accuracy of 88.22% [26]. This method has a good computational accuracy and produces high classification accuracy, but the performance depends on the classifier used.

Hybrid ReliefF and Rough Set (RFRS) [27] extracts features utilizing the ReliefF algorithm and reduces features using the Rough Set heuristic reduction algorithm. In the experiment using the Statlog heart dataset, RFRS feature selection with ensemble classifier produced an accuracy of 92.59% [27]. The weakness of this method is that the parameter for the ReliefF algorithm is unstable. This method can take the benefits of the two feature selection approaches to produce the best results. The advantages of this method include removing redundant features effectively, accelerating the Rough Set reduction process, reducing the number of reductions, and improving reduction quality.

A combination of SVM-RFE and gain-ratio [28] is used to improve accuracy and reduce computing time. The SVM-RFE

TABLE I
SUMMARY OF FEATURE SELECTION APPROACHES FOR PREDICTION OF HEART DISEASE

Approach	Merits	Demerits
Filter	<ul style="list-style-type: none"> • Computationally efficient • Computationally cheap • Scalable to a large dataset • Independent • Low complexity 	<ul style="list-style-type: none"> • Ignore feature dependencies • Ignore interaction with classifier
Wrapper	<ul style="list-style-type: none"> • Better performance accuracy than filter • Interact with classifier • Consider feature dependencies 	<ul style="list-style-type: none"> • Computationally expensive • Computationally complex • Prone to over-fitting in the small training dataset
Embedded	<ul style="list-style-type: none"> • Interact with classifier • Consider feature dependencies • Computational complexity is better than a wrapper • Computationally inexpensive than a wrapper • Less over-fitting compared to a wrapper 	<ul style="list-style-type: none"> • Classifier specific • Computationally costlier than filter
Hybrid	<ul style="list-style-type: none"> • Performance accuracy is better than filter • Less computational complexity than a wrapper • Less over-fitting than a wrapper 	<ul style="list-style-type: none"> • Classifier specific

results and the gain-ratio are combined to calculate the final feature weights. Features with a weight above the threshold are selected as the last feature set. In the research of predicting heart disease, combination SVM-RFE and gain ratio and Naïve Bayes classifier achieved optimal results with an accuracy of 84.1584% when using ten features [28]. Random Forest classifier yielded an accuracy of 84.1604% with 12 features [28]. This method can decrease the complexity of the model and improve prediction accuracy.

The summary of feature selection approaches for the prediction of heart disease, the merits and demerits of the approaches can be seen in Table I.

III. CLASSIFICATION APPROACHES FOR THE PREDICTION OF HEART DISEASE

Classification is a data mining task that is often used in healthcare, one of which is to predict disease [29]. There are two steps in classification; the first step is the learning step where the model is built and trained using training data that has

been determined with class labels, and the second step is the classification in which the model is used to predict class labels for test data to estimate the accuracy of the classifier model [30]. Many researchers have conducted a study on predicting heart disease using various classification approaches, including traditional machine learning, ensemble learning, hybrid approach, and deep learning.

A. Traditional Machine Learning

Traditional machine learning learns from the data using a specific algorithm to make a prediction. Several traditional machine learning techniques have been applied for predicting heart disease, such as decision tree, neural network, Support Vector Machine (SVM), naïve Bayes, and K-Nearest Neighbor (K-NN) [26], [31]-[34]. The traditional machine learning techniques work well in heart disease prediction. However, each technique also has its shortcomings.

A decision tree was implemented on heart disease prediction [31]. The decision tree builds classification models in a tree structure. A decision tree divides the dataset into subsets, and at the same time, the related decision tree is built gradually. In the experiment of heart disease prediction, the decision tree yielded an accuracy of 98.28% [31]. A decision tree is computationally inexpensive; it can quickly construct a model even on a large training set. This method is easy to understand but unstable because slight changes to the data will have a large effect on the decision tree structure.

A neural network was used on heart disease prediction [32]. The methodology selects the important variables related to heart disease, then builds a neural network model using the training dataset and uses the model to predict heart disease using the testing dataset. Using a neural network with the selected features generated by logistic regression produced accuracy, sensitivity, and specificity of 84%, 91.4%, 77.5%, respectively [32]. The limitations of this method are the black-box problem, which makes it difficult to explain where the output comes from, and it is prone to over-fitting. The advantages include needing less formal statistical training to develop, detecting all possible interactions between predictor variables, and discovering complex relationships between dependent and independent variables.

Support Vector Machine (SVM) was applied in predicting heart disease [26]. SVM can generate an optimal hyperplane to classify new instances using labeled training data. The experiment on the Cleveland heart disease dataset using SVM with selected features generated by PSO-SVM yielded an accuracy of 88.22% [26]. SVM is a stable model that a small change of the data does not really affect the hyperplane. It also provides flexibility using the concept of kernels and has a good generalization. However, it is not suitable for a large dataset; performance decreases when there is a lot of noise and difficult to interpret.

Naïve Bayes was implemented in the prediction system for heart disease [33]. Naïve Bayes is a classifier that has an independent assumption between features. On the Statlog heart dataset, naïve Bayes using features selected by Particle Swarm Optimization (PSO) achieved an accuracy of 87.91% [33]. This method is simple, fast, and highly scalable. The limitation of

this method is it assumes that the attributes are independent, so there is a need for measuring the correlations among features before using Naive Bayes.

K-Nearest Neighbor (K-NN) was implemented on the heart disease prediction system [34]. In the research, a correlation matrix was used to select the features, then training the model using a training dataset, and the prediction was performed using a test dataset. On the UCI heart disease dataset, K-NN produced an accuracy of 88.52% and sensitivity of 91.17% [34], which was higher than other classifiers, such as SVM, naïve bayes, and decision tree. This method is simple and robust to noise. However, it degrades the performance on a large dataset and requires large memory.

B. Ensemble Learning

Ensemble Learning is the aggregation of the prediction of multiple classifiers [35]. The aggregate result of multiple classifiers is less noisy than the individual classifier, which makes it more stable and robust. Ensemble improves the efficacy of the weak learners by combining it with strong learners. The ensemble learning method can produce better prediction compared to a single classifier [36]. It also can avoid over-fitting. However, it increases the complexity leading to a lack of interpretability. It is also computationally expensive.

SVM with boosting technique was implemented to predict heart disease and outperformed other methods [35]. The ensemble method used was boosting applied on the SVM classifier. The method repeatedly ran on training data, then allowed the learned classifier to vote. On the Cleveland heart disease dataset, SVM with boosting technique achieved an accuracy of 84.81% [35]. Ensemble boosting can decrease the bias error. However, it tends to over-fitting and sensitive to outliers.

Ensemble majority vote was applied on a prediction of heart disease and compared to other ensemble methods [36]. The Majority vote merges multiple classifiers by majority voting. From the Cleveland dataset experiment, the ensemble majority vote method with Naive Bayes, Bayes Net, Random Forest, and Multilayer Perceptron with the chosen features achieved an accuracy of 85.48%, higher than other ensemble methods [36]. The Majority vote can improve the accuracy of weak classifiers, but it is computationally expensive.

Ensemble-based on distances for the K-NN method was proposed to predict heart disease [37]. In the experiments, the ensemble method implemented using three distances and five distances. Then, they added a weighted version based on the average accuracy of each distance using the K-NN method. The weighted 3-distance ensemble produced an accuracy of 84.83%, which was the highest of other configurations [37]. This method is easy to implement, but there is a need to evaluate the optimal configuration for classification.

Random forest is an ensemble that develops forests from some decision trees [38]. Classification is done by voting from each tree; then random forest chose the class that got more votes. On the Cleveland dataset, random forest produced accuracy, sensitivity, and specificity of 87.50%, 86.67%, 88.37%, which were higher than other methods [38]. The advantages of this method include the good performance

accuracy and the ability to avoid an over-fitting problem. The limitation of this method is that it has a slower computation.

C. Hybrid Approach

The hybrid approach generates a predictive model by merging two or more techniques [39]. The hybrid model can reduce the limitation of individual models and exploit their generalization methods [40]. Although the hybrid approach also uses information fusion concept like an ensemble, it has a slightly different way. The ensemble combines several homogeneous weak models, whereas hybrid combines heterogeneous machine learning approaches looking for a homogeneous solution [41]. Using a hybrid approach to predict heart disease can improve the predictive performance, but it depends on the combination of the techniques.

A Hybrid Classifier with Weighted Voting (HCWV) was proposed for heart disease prediction [42]. The method employed nine classifiers to be trained on training data and evaluated the accuracy of each model on testing data. The models were arranged in ascending order based on the accuracy, then combined for ensemble into several combinations. The combinations were then ensembled again to predict the final output. This method achieved an accuracy of 82.54% on the UCI heart disease dataset [42]. This method shows a good performance accuracy compared to single classifiers and ensemble classifiers. However, this method has more computation complexity because it uses a combination of multiple methods.

A combination of Variable Precision Rough Set (VPRS) and Repeated Incremental Pruning Error Reduction (RIPPER) was implemented for detecting coronary heart disease [3]. The Cleveland dataset experiment using combined two rule mining methods, namely VPRS and RIPPER, with the features generated by VPRS feature selection resulted in accuracy, sensitivity, and specificity of 88.89%, 100%, 77.08%, respectively [3]. The combination of VPRS and RIPPER has better performance results than that of the individually separated method. This method can improve the quality of reasoning and increase performance, but this method is also increasing the complexity of the algorithm.

The hybrid ensemble method was proposed for the prediction of heart disease [43]. This method combines initial features of samples and base classifier predictions to generate a new feature vector for the fuser classifier. The hybrid ensemble method produced accuracy, sensitivity, and specificity of 96%, 80%, 93% [43] on the SPECT heart disease dataset. The hybrid ensemble method produces better performance than the basic ensemble method. This method's limitation is that it depends on the classifier used as the base classifier and fuser classifier.

Hybrid Random Forest with Linear Model (HRFLM) was implemented to predict heart disease [39]. The method combined the Random Forest and Linear method characteristics and used feature selection based on decision tree entropy to predict heart disease. HRFLM produced accuracy, sensitivity, and specificity of 88.47%, 92.8%, 82.6% [39] on the Cleveland dataset. This method improves the performance accuracy and has low classification error. However, this method also increases computational complexity.

D. Deep Learning

Deep Learning is a branch of machine learning that consists of high-level abstraction modelling algorithms on data using multiple processing layers [44]. The advantage of deep learning over traditional machine learning is that it avoids the need for feature engineering processes [45]. It has the capacity to execute feature engineering on its own. It also performs very well on a voluminous dataset. However, deep learning needs a lot of examples to learn a concept [46]. It is also computationally more expensive than traditional machine learning.

Deep Neural Network (DNN) was implemented for the prediction of heart disease [47]. DNN uses the neural network concept with several layers, and each layer has several neurons. The model has four layers that consisted of an input layer, two hidden layers, and an output layer. On the Cleveland heart disease dataset, the deep neural network yielded an accuracy of 83.67% [47]. This method has the ability to execute the feature engineering by itself but computationally expensive.

Recurrent Neural Network (RNN) was implemented in the prediction of coronary heart disease [48]. RNN is a neural network where connections between hubs form a coordinated chart along with a temporal sequence. RNN can utilize internal memory to process sequences of inputs. RNN can handle large data and overcome delay and noise. RNN got an accuracy of more than 90% in coronary heart disease prediction using the dataset from the local city of the United States of America [48]. This method can handle arbitrary inputs and output lengths, remember every information through time, and utilize internal memory to process the arbitrary series of inputs. The limitations of this method include slow computation and prone to exploding and gradient vanishing problems.

Convolutional Neural Network (CNN) was used in heart disease prediction [49]. The method consists of a convolutional layer, pooling layer, and fully connected layer. The convolutional layers convolve the input, the pooling layer reduces the dimensionality of input data, and the fully connected layer links every neuron in one layer to another layer. ReLu is used as an activation function in hidden layers, and sigmoid is used in the output layer for classification. In the research, CNN resulted in an accuracy of 85% for heart disease prediction [49]. It is higher than other algorithms, such as naïve Bayes and K-NN. This method automatically detects the important features and can be used with both structured and unstructured data. However, this method requires a lot of data and computationally expensive.

The summary of the literature survey of classification and feature selection approaches for predicting heart disease and their performance accuracy for a given dataset can be seen in Table II at the end of this paper.

IV. CONCLUSION

Feature selection can be used to choose the most relevant features and improve the performance of the predictive performance. Feature selection approaches include filter, wrapper, embedded, and hybrid method. A filter is computationally faster than other methods. Wrapper and

embedded can find optimal features that have dependencies on each other and interact with the classifier. The hybrid method combines several feature selection methods to get optimal results by exploiting each method's advantages.

Classification is a data mining task that can be utilized to predict heart disease. Classification approaches include traditional machine learning, ensemble learning, hybrid approach, and deep learning. Traditional machine learning is a fundamental method that uses a specific algorithm to learn from the data and make a prediction. The ensemble learning aggregates the prediction of several classifiers to improve the performance of a single classifier. The hybrid approach combines two or more techniques to reduce individual models' limitation and exploit their generalization methods. Deep learning is an efficient method because it does not require a predetermined feature engineering process.

As there is high mortality due to heart disease, early detection of heart disease is very helpful for patients and health professionals. Many researchers have been utilizing the development of information technology to develop a heart disease computer-assisted diagnosis system. This study contributes by reviewing the feature selection and classification approaches in predicting heart disease, which can be a consideration for future research in developing a heart disease computer-assisted diagnosis system. 'xx'

REFERENCES

- [1] A. Rodgers, C.M.M. Lawes, T. Gaziano, and T. Vos, "The Growing Burden of Risk from High Blood Pressure, Cholesterol, and Bodyweight," in *Disease Control Priorities in Developing Countries*. 2nd ed., D.T. Jamison, J.G. Breman, A.R. Measham, G. Alleyne, M. Claeson, D.B. Evans, P. Jha, A. Mills, and P. Musgrove, Eds., New York, USA: Oxford University Press, 2006, pp. 851-868.
- [2] Y. Yan, J.W. Zhang, G.Y. Zang, and J. Pu, "The Primary use of Artificial Intelligence in Cardiovascular Diseases: What Kind of Potential Role Does Artificial Intelligence Play in Future Medicine?," *Journal of Geriatric Cardiology*, Vol. 16, No. 2019, pp. 585-591, 2019.
- [3] D. Normawati and S. Winarti, "Feature Selection with Combination Classifier Use Rules-Based Data Mining for Diagnosis of Coronary Heart Disease," *12th International Conference on Telecommunication Systems, Services, and Applications (TSSA)*, 2018, pp. 1-6.
- [4] J. Li, K. Cheng, S. Wang, F. Morstatter, R.P. Trevino, J. Tang, and H. Liu, "Feature Selection: A Data Perspective," *ACM Computing Surveys*, Vol. 50, No. 6, pp.1-45, 2017.
- [5] N. Hoque, M. Singh, and D.K. Bhattacharyya, "EFS-MI: An Ensemble Feature Selection Method for Classification," *Complex and Intelligent System*, Vol. 4, pp. 105-118, 2018.
- [6] G. Kesavaraj and S. Sukumaran, "A Study on Classification Techniques in Data Mining," *2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, 2013, pp. 1-7.
- [7] D. Jain and V. Singh, "Feature Selection and Classification Systems for Chronic Disease Prediction: A Review," *Egyptian Informatics Journal*, Vol. 19, No. 3, pp. 179-189, 2018.
- [8] I. Guyon and A. Elisseeff, "An Introduction of Variable and Feature Selection," *Journal of Machine Learning Research*, Vol. 3, pp. 1157-1182, 2003.
- [9] G. Chandrashekar and F. Sahin, "A Survey on Feature Selection Methods," *Computers and Electrical Engineering*, Vol. 40, No. 1, pp. 16-28, 2014.
- [10] S. Visalakshi and V. Radha, "A Literature Review of Feature Selection Techniques and Applications: Review of Feature Selection in Data Mining," *2014 IEEE International Conference on Computational Intelligence and Computing Research*, Coimbatore, 2014, pp. 1-6.
- [11] M.W. Mwadulo, "A Review on Feature Selection Methods for Classification Tasks," *International Journal of Computer Applications Technology and Research*, Vol. 5, No. 6, pp. 395-402.
- [12] J.C. Ang, A. Mirzal, H. Haron, and H.N.A. Hamed, "Supervised, Unsupervised, and Semi-Supervised Feature Selection: A Review on Gene Selection," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, Vol. 13, No. 5, pp. 971-989, 2016.
- [13] A. Khempila and V. Boonjing, "Heart Disease Classification Using Neural Network and Feature Selection," *21st International Conference on Systems Engineering*, 2011, pp. 406-409.
- [14] H. Takci, "Improvement of Heart Attack Prediction by the Feature Selection Methods," *Turkish Journal of Electrical Engineering & Computer Sciences*, Vol. 26, No. 1, pp. 1-10, 2018.
- [15] S. Bashir, Z.S. Khan, F.H. Khan, A. Anjum, and K. Bashir, "Improving Heart Disease Prediction Using Feature," *16th International Bhurban Conference on Applied Sciences & Technology (IBCAST)*, 2019, pp. 619-623.
- [16] A.M. Usman, U.K. Yusof, and S. Naim, "Cuckoo Inspired Algorithms for Feature Selection in Heart," *International Journal of Advances in Intelligent Informatics*, Vol. 4, No. 2, pp. 95-106, 2018.
- [17] D.A.A.G. Singh, S.A. Balamurugan, and E.J. Leavline, "Literature Review on Feature Selection Methods for High-Dimensional Data," *International Journal of Computer Applications*, Vol. 136, No. 1, pp. 9-17, 2016.
- [18] R.S. El-Sayed, "Linear Discriminant Analysis for an Efficient Diagnosis of Heart Disease via Attribute Filtering Based on Genetic Algorithm," *Journal of Computers*, Vol. 13, No. 11, pp. 1290-1299, 2018.
- [19] J. Zeniarja, A. Ukhifahdhina, and A. Salam, "Diagnosis of Heart Disease Using K-Nearest Neighbor Method Based on Forward Selection," *Journal of Applied Intelligent System*, Vol. 4, No. 2, pp. 39-47, 2019.
- [20] A.U. Haq, J. Li, M.H. Memon, M.H. Memon, J. Khan, and S.M. Marium, "Heart Disease Prediction System Using Model of Machine Learning and Sequential Backward Selection Algorithm for Features Selection," *5th International Conference for Convergence in Technology (I2CT)*, 2019, pp. 1-4.
- [21] U.M. Khaire and R. Dhanalakshmi, "Stability of Feature Selection Algorithm: A Review," *Journal of King Saud University – Computer and Information Sciences*, pp. 1-14, 2019, <https://doi.org/10.1016/j.jksuci.2019.06.012>.
- [22] K. Rajeswari, V. Vaithiyanathan, and T.R. Neelakantan, "Feature Selection in Ischemic Heart Disease Identification using Feed Forward Neural Networks," *Procedia Engineering*, Vol. 41, pp. 1818-1823, 2012.
- [23] V. Sabarinathan and V. Sugumaran, "Diagnosis of Heart Disease Using Decision Tree," *International Journal of Research in Computer Applications & Information Technology*, Vol. 2, No. 6, pp. 74-79, 2014.
- [24] C. Yang, B. An, and S. Yin, "Heart-Disease Diagnosis via Support Vector Machine Based Approaches," *IEEE International Conference on Systems, Man, and Cybernetics*, 2018, pp. 3153-3158.
- [25] H.-H. Hsu, C.-W. Hsieh, and M.-D. Lu, "Hybrid Feature Selection by Combining Filters and Wrappers," *Expert Systems with Applications*, Vol. 38, No. 7, pp. 8144-8150, 2011.
- [26] J. Vijayashree and H.P. Sultana, "A Machine Learning Framework for Feature Selection in Heart Disease Classification Using Improved Particle Swarm Optimization with Support Vector Machine Classifier," *Program Comput Software*, Vol. 44, p. 388-397, 2018.
- [27] X. Liu, X. Wang, Q. Su, M. Zhang, Y. Zhu, Q. Wang, and Q. Wang, "A Hybrid Classification System for Heart Disease Diagnosis Based on the RFRS Method," *Computational and Mathematical Methods in Medicine*, Vol. 2017, pp. 1-11, 2017.

- [28] K. Pahwa and R. Kumar, "Prediction of Heart Disease Using Hybrid Technique for Selecting Features," *4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer, and Electronics (UPCON)*, 2017, pp. 500-504.
- [29] S.C. Pandey, "Data Mining Techniques for Medical Data: A Review," *International Conference on Signal Processing, Communication, Power and Embedded System (SCOPE)-2016*, 2016, pp. 972-982.
- [30] A.H. Seh and P.K. Chaurasia, "A Review on Heart Disease Prediction Using Machine Learning Techniques," *International Journal of Management, IT & Engineering*, Vol. 9, No. 4, pp. 208-224, 2019.
- [31] A.S. Karthiga, M.S. Mary, and M. Yogasini, "Early Prediction of Heart Disease Using Decision Tree Algorithm," *International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)*, Vol. 3, No. 3, pp. 1-16, 2017.
- [32] K.U. Maheswari and J. Jasmie, "Neural Network-based Heart Disease Prediction," *International Journal of Engineering Research & Technology (IJERT)*, Vol. 5, No. 17, pp. 1-4, 2017.
- [33] U.N. Dulhare, "Prediction System for Heart Disease Using Naive Bayes and Particle Swarm Optimization," *Biomedical Research 2018*, Vol. 29, No. 12, pp. 2646-2649, 2018.
- [34] A. Gupta, S. Yadav, S. Shahid, and V. U, "HeartCare: IoT based Heart Disease Prediction System," *2019 International Conference on Information Technology (ICIT)*, 2019, pp. 88-93.
- [35] S. Pouriyeh, S. Vahid, G. Sannino, D.G. Pietro, H. Arabnia, and J. Gutierrez, "A Comprehensive Investigation and Comparison of Machine Learning Techniques in the Domain of Heart Disease," *2017 IEEE Symposium on Computers and Communications (ISCC)*, 2017, pp. 1-4.
- [36] C.B.C. Latha and S.C. Jeeva, "Improving the Accuracy of Prediction of Heart Disease Risk Based on Ensemble Classification Techniques," *Informatics in Medicine Unlocked*, Vol. 16, pp. 1-9, 2019.
- [37] A.P. Pawlovsky, "An Ensemble Based on Distances for a kNN Method for Heart Disease Diagnosis," *2018 International Conference on Electronics, Information, and Communication (ICEIC)*, 2018, pp. 1-4.
- [38] A. Ed-daoudy and K. Maalmi, "Performance Evaluation of Machine Learning based Big Data Processing Framework for Prediction of Heart Disease," *2019 International Conference on Intelligent Systems and Advanced Computing Sciences (ISACS)*, 2019, pp. 1-5.
- [39] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," *IEEE Access*, Vol. 7, p. 81542–81554, 2019.
- [40] V. Miškovic, "Machine Learning of Hybrid Classification Models for Decision Support," *SINTEZA 2014*, 2014, pp. 318-323.
- [41] P. Kazienco, E. Lughofer, and B. Trawinski, "Editorial on the Special Issue 'Hybrid and Ensemble Techniques in Soft Computing: Recent Advances and Emerging Trends'," *Soft Computing*, Vol. 19, p. 3353–3355, 2015.
- [42] M. Saini, N. Baliyan, and V. Bassi, "Prediction of Heart Disease Severity with Hybrid Data Mining," *2017 2nd International Conference on Telecommunication and Networks (TEL-NET 2017)*, 2017, pp. 1-6.
- [43] E. Nikookar and E. Naderi, "Hybrid Ensemble Framework for Heart Disease Detection and Prediction," *(IJACSA) International Journal of Advanced Computer Science and Applications*, Vol. 9, No. 5, pp. 243-248, 2018.
- [44] B.-B. Benuwa, Y. Zhan, B. Ghansah, D.K. Wornyo, and F.B. Kataka, "A Review of Deep Machine Learning," *International Journal of Engineering Research in Africa*, Vol. 24, pp. 124-136, 2016.
- [45] A. Hernández-Blanco, B. Herrera-Flores, D. Tomás and B. Navarro-Colorado, "A Systematic Review of Deep Learning Approaches to Educational Data Mining," *Complexity*, Vol. 2019, pp. 1-22, 2019.
- [46] B. Zohuri and M. Moghaddam, "Deep Learning Limitations and Flaws," *Modern Approaches on Material Science*, pp. 241-250, 2020.
- [47] K.H. Miao and J.H. Miao, "Coronary Heart Disease Diagnosis using Deep Neural Networks," *(IJACSA) International Journal of Advanced Computer Science and Applications*, Vol. 9, No. 10, pp. 1-8, 2018.
- [48] P. Anandajayam, C. Krishnakoumar, S. Vikneshvaran, and B. Suryanarayanan, "Coronary Heart Disease Predictive Decision Scheme Using Big Data and RNN," *Proceeding of International Conference on Systems Computation Automation and Networking 2019*, 2019, pp. 1-6.
- [49] V. Shankar, V. Kumar, U. Devagade, V. Karanth, and K. Rohitaksha, "Heart Disease Prediction Using CNN Algorithm," *SN Computer Science*, Vol. 1, pp. 1-8, 2020.

TABLE II
SUMMARY OF LITERATURE SURVEY

Classification	Feature Selection	Dataset	Accuracy	Ref.
Neural Network	Information gain	Cleveland	Training: 89.56%, Validation: 80.99%	[13]
SVM	ReliefF	Statlog	84.81%	[14]
SVM	MRMR	UCI	84.85%	[15]
SVM	Filter-based Cuckoo Search Algorithm (CSA)	Eric	89.90%	[16]
		Hungarian	94.22%	
		Statlog	94%	
		Z-Alizadeh Sani	90%	
Naïve Bayes		Echocardiogram	100%	
Linear Discriminant Analysis	Genetic algorithm	Cleveland	89.07% for binary class, 67.22% for multiclass	[18]
K-NN	Forward selection	NM*	78.66%	[19]
K-NN	SBS	Cleveland	90%	[20]
Neural Network	ANN	Ischemic Heart Disease dataset of Madras Medical College	Training: 89.4%, Testing: 82.2 %	[22]

TABLE II (CONT'D)
SUMMARY OF LITERATURE SURVEY

Classification	Feature Selection	Dataset	Accuracy	Ref.
Decision tree	Gain ratio decision tree	NM*	85%	[23]
PCA-SVM	SVM-RFE	UCI	88.24%	[24]
SVM	PSO-SVM	Cleveland	88.22%	[26]
Ensemble	Hybrid ReliefF and Rough Set	Statlog	92.59%	[27]
Naïve Bayes	SVM-RFE and gain-ratio	Cleveland	84.1584%	[28]
Random Forest			84.1604%	
Decision tree	-	UCI	98.28%	[31]
Neural Network	Logistic regression	Cleveland	84%	[32]
Naïve Bayes	PSO	Statlog	87.91%	[33]
K-NN	Correlation matrix	UCI	88.52%	[34]
SVM with boosting	-	Cleveland	84.81%	[35]
Ensemble majority vote	Brute force	Cleveland	85.48%	[36]
Ensemble based on distances for K-NN	-	Cleveland	84.83%	[37]
Random forest	-	Cleveland	87.50%	[38]
Hybrid Classifier with Weighted Voting	-	UCI	82.54%	[42]
VPRS+RIPPER	VPRS	Cleveland	88.89%	[3]
Hybrid ensemble	-	SPECT	96%	[43]
Hybrid Random Forest with Linear Model	Decision Tree Entropy	Cleveland	88.47%	[39]
Deep Neural Network	-	Cleveland	83.67%	[47]
Recurrent Neural Network	-	Local dataset	90%	[48]
Convolutional Neural Network	-	NM*	85%	[49]

NM*: Not mentioned in the research paper clearly.