# Prediction of Heart Diseases Using Majority Voting Ensemble Method

Dželila Mehanović, Zerina Mašetić, and Dino Kečo

## Abstract

Heart disease is the one of the most serious problems in healthcare and affects large number of people. It is very important to detect it on time, otherwise it can cause serious consequences, such as death. In this paper, we applied artificial neural network, k-nearest neighbor and support vector machine algorithms to build model which will be used for prediction of heart disease. Multiple experiments using each of these algorithms are performed. Additionally, the ensemble learning is applied, and results are compared. Initially, the problem was approached as multiclass classification, however it was transformed into the binary classification problem, to simplify model since number of outputs is reduced from five to two. In both cases, the highest accuracies are obtained by majority voting which are 61.16% for multiclass classification and 87.37% for binary classification.

## Keywords

Heart disease • Classification • ANN • SVM • KNN • Ensemble learning

## 1 Introduction

Since healthcare organizations are complex systems they generate vast amount of data that comes in different formats. Because of this, the key challenge is to build intelligent systems that will efficiently interpret generated data and support humans in decision making [1, 2].

To build such system, different solutions such as artificial intelligence methods, linked data, semantic web technologies and NoSQL datastores are combined. In recent years, machine learning becomes popular for developing intelligent systems in healthcare. Machine learning algorithms are capable to approximate relationships between dataset variables in form of a function that is used for prediction and decision making. Application of machine learning models in healthcare systems improve efficiency and accuracy of the system overall [3].

Google employed machine learning to detect breast cancer by detecting patterns in the tissue with 87% accuracy, which is better than 73% accuracy achieved by human [4]. Scientists from Stanford developed algorithm for skin cancer detection using visual processing and deep learning inspired by neural networks [5]. In the [6] algorithm for detection of diabetic retinopathy using neural networks is presented.

One of the most common problems in healthcare is heart disease. Based on the latest statistics, number of people with heart disease is predicted to rise by 46% by 2030, which is more than 8 million adults with heart disease. According to World Health Organization [7], cardiovascular diseases (a group of disorders of the heart and blood vessels) are number one cause of death all over the world [8]. Statistical report of American Heart Association for 2017 [9] shows that about 92.1 million Americans are living with some cardiovascular disease.

Gi Beom Kim, in his paper [10] presents that according to database of Korean Health Foundation, about 50,000 of adults with heart disease live in South Korea and more than 4000 enter adulthood every year. At the current rate, it is estimated that about 70,000 of adults with heart disease will live there by the year 2020. Approximately 2200 people die of cardiovascular disease each day, which is one death on average every 40 s.

In recent researches [12–16] various machine learning algorithms have been applied to predict heart disease problems. Different techniques are employed to build reliable system which will produce useful results while lowering costs and diagnosis time. The goal of this paper is to build a model that will combine multiple classification algorithms to predict heart disease and to compare single algorithm models

D. Mehanović (✉) · Z. Mašetić · D. Kečo
International Burch University, Sarajevo 71000,
Bosnia and Herzegovina
e-mail: dzelila.mehanovic@stu.ibu.edu.ba

and ensemble model. Majority voting has been used as ensemble method for combining multiple machine learning algorithms [11].

The rest of paper is organized as follows: in Sect. 2 we summarize related works, in Sect. 3 we describe dataset and applied methods, in Sect. 4 we present results of our work and finally, we conclude our work with Sect. 5.

## 2 Literature Review

In the March 2017, Singh et al. [12] proposed web application that enables users to share their heart related problems and get diagnosis of disease using intelligent system online. Application takes inputs from user, process them and returns disease related to inputs from user. To avoid variance, dataset with 14 input attributes was split indiscriminately into two sets. Finally, implementation of model was performed using Naïve Bayes' classifier. As result, implemented application returns output to users based on prediction whether risk for heart disease is low, average or high.

Various data mining techniques were applied to predict heart disease. Devi et al. [13] analysed classification techniques for decision making in this field, especially Decision trees, Naive Bayes, Neural Networks and Support Vector Machines. They found out that application of hybrid data mining techniques can give promising results. Combining the outputs of each algorithm and comparing them helps to make prediction quicker and more accurate.

Datasets related to same disease problem may show different results applying same machine learning techniques. El-Bialy et al. [14] focused on integration of results of machine learning techniques applied on sets for heart diseases. They applied fast decision tree and C4.5 tree techniques, after which they compared features in trees resulted from different datasets. Common features among these datasets are collected to create new dataset which is used in later analysis. It is shown that accuracy of new dataset is higher than average accuracy of all separate datasets. Average accuracy of all datasets was 75.48% using fast decision tree and 76.30% using C4.5. Classification accuracy for newly collected dataset was 78.06 and 77.50% for fast decision tree and C4.5, respectively.

Venkatalakshmi and Shivsankar [15] presented prediction system for heart disease based on predictive mining. Experiments were carried out using Weka, open source tool for data mining, and data from UCI Machine Learning Repository. The goal was to compare performance of predictive data mining techniques such as Naive Bayes and Decision tree. Naive Bayes outperformed Decision Tree with accuracy of 85.03, while accuracy of Decision tree was 84.01.

In their research paper, Jabbar et al. [16] combined K-Nearest Neighbor and Genetic Algorithm on seven datasets to build heart disease classifier. The results of their study showed that accuracy is increased 5% using both, KNN and GA, rather than only KNN. Also, accuracy is decreasing as k-value goes on increasing. Although emphasis was on data related to Andhra Pradesh, city in India, it is shown that classifier gives high accuracy when it is applied on other heart disease datasets.

Table 1 shows summary of experiments mentioned above. It displays accuracy of heart disease classifiers obtained by several authors applying various methods.

Authors in [14, 15] used same dataset as we did, but unlike them, additionally, we want to apply multiclass classification to evaluate results. Besides that, our main goal is to explore gain achieved by application of ensemble learning.

## 3 Methodology

### 3.1 Dataset

As a part of this research, multiple machine learning models were developed using Heart Disease dataset from UCI Machine Learning Repository [17]. Original dataset contains 76 attributes, but all published works used only 14 of them, so we did same in our work. These attributes are selected as the most important for the reliable prediction. Data set contains 303 instances and it is publicly available.

The last attribute in dataset represents diagnosis of heart disease. Value 0 indicates absence of heart disease and values 1, 2, 3 and 4 indicate different levels of disease. Analyzing representation of each class individually, which are 54.12, 18.15, 11.88, 11.55 and 4.29% respectively, we can conclude that this dataset does not have skewed class problem.

Figure 1 shows visual representation of all attributes in dataset. It presents distribution of attributes in respect to particular class in dataset.

To split dataset, we applied two approaches. Firstly, we splitted data into training and testing set with ratio 66:34. Later, we applied 10-fold cross validation to compare results.

### 3.2 Feature Selection

After dataset was uploaded, we performed feature selection process which is used to identify attributes that are most relevant for prediction [18]. Two elevators were tried: GainRatioAttributeEval and InfoGainAttributeEval.

**Table 1** Accuracy of various methods for heart disease classification in related work

| # | Authors | Methods | Results |
|---|---------|---------|---------|
| 1 | El-Bialy et al. [14] | Fast decision tree<br>C4.5 | 78.06%<br>77.50% |
| 2 | Venkatalakshmi and Shivsankar [15] | Naive Bayes<br>Decision tree | 85.03%<br>84.01% |
| 3 | Jabbar et al. [16] | K-nearest neighbor and genetic algorithm | Accuracy is increased by 5% using both, KNN and GA, rather than only KNN |



**Fig. 1** Visual representation of attributes

According to the output of these two evaluators, several attributes were excluded from the further examination, as they showed no significant impact on an automated heart disease prediction process. Excluded attributes are resting blood pressure, serum cholesterol, fasting blood sugar and resting electrocardiographic result.

### 3.3 Classification

In our work, we applied ensemble learning to build model. Ensemble learning is used to combine several models to improve results. Multiple methods of ensemble learning exist such as voting, stacking, bagging and boosting which are explained at [11] in detail. In our work we investigate majority voting. Each model included in majority voting makes its own prediction and final prediction is the one with highest number of votes.

We considered three classification algorithms: artificial neural network (ANN), support vector machine (SVM) and k-nearest neighbors algorithm (KNN). These algorithms are combined together and complement each other. Figure 2 shows the process of implementation.

Different combinations of parameters were used for each of these algorithms. Final model was built using the parameter values as provided in continuation.

**Fig. 2** Process of model
implementation
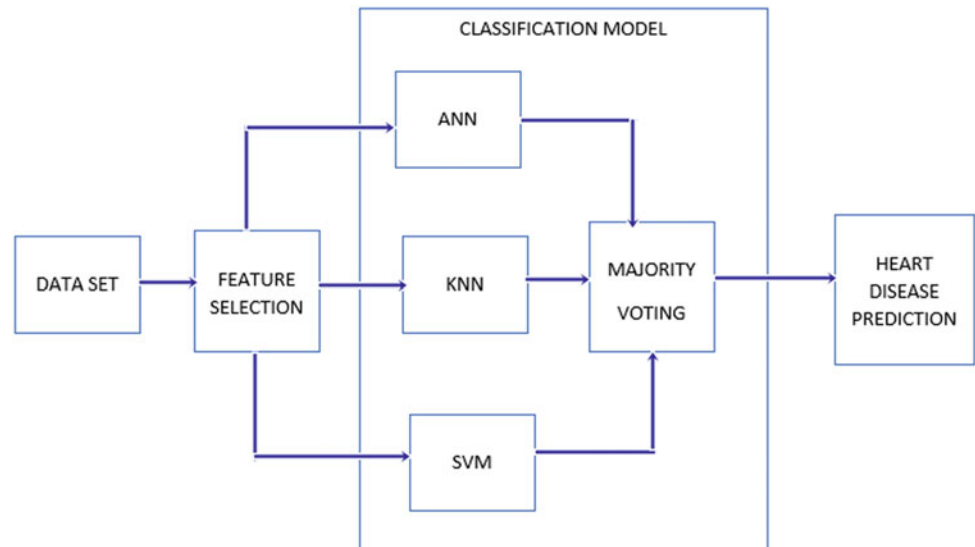


Artificial neural network contained 1 hidden layer and 0.4 learning rate. The value of training time was increased from default value of 500 to value of 5000. K-nearest neighbor was applied with value of k equal to 5. Also, we applied LinearNNSearch with Filtered Distance. Support vector machine was used with polynomial kernel and value of exponent equal to 1.5. The regularization factor was 3.0.

Described parameters were selected as the best combination for these algorithms according to obtained accuracy values.

Initial dataset had 5 output classes (class 0 for absence of disease and classes 1, 2, 3, 4 for presence) which are used for multiclass classification. Moreover, to compare results we applied binary classification with two classes (class 0 for absence and class 1 for presence of disease). Finally, for each of them we tested performance of ensemble learning.

## 4   Results

To measure performance of classifiers, we calculated the accuracy, the specificity and the precision.

The accuracy represents ratio of correctly classified samples and total number of samples [19].

$$\text{ACC} = \frac{\text{TP} + \text{FN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \qquad (1)$$

The sensitivity or true positive rate (TPR) is ratio of true positives and actual positives (TP + FN) [19].

$$\text{TPR} = \frac{TP}{TP + FN} \qquad (2)$$

The precision or positive predictive value (PPV) is ratio of true positives and predicted positives [19].

$$\text{PPV} = \frac{TP}{TP + FP} \qquad (3)$$

From confusion matrix presented in Table 3, we can see types of mismatching between classes and notice that errors mostly occur between neighboring classes.

Table 2 and Fig. 3 show results obtained by multiclass classification and percentage split. The highest accuracy 61.16 is gained by majority voting, while ANN and KNN resulted in same accuracy 58.25, when applied separately.

From confusion matrix presented in Table 4, we can see that mismatch occurs between classes 0, 1, 2 and 3, not only between neighboring classes. But errors are reduced by increasing the distance between neighbored classes.

Table 3 and Fig. 4 show results obtained by multiclass classification and 10-fold cross validation. Again, majority voting achieved highest accuracy 58.41. In this case KNN outperformed ANN, while SVM still had the lowest accuracy.

If we compare results obtained by percentage split and cross validation for multiclass classification, we can observe that for the most of measurements percentage split results are higher than those of cross validation. Only ANN resulted in higher accuracy by cross validation than by percentage split.

In the next experiment, problem has been transformed into binary classification with output labels 0 and 1 where 0 presents absence and 1 presence of heart disease.

Table 4 and Fig. 5 show results obtained by binary classification and percentage split. It can be noticed that results are higher than those obtained by multiclass classification. Majority voting resulted in highest accuracy 87.37. Moreover, contrary to results of multiclass classification, KNN obtained lowest accuracy and SVM has highest accuracy, when applied without ensemble.

**Table 2** Multi class classification results by percentage split

| Algorithm | Sensitivity | Precision | Accuracy |
|---|---|---|---|
| ANN | 0.58 | 0.50 | 58.25 |
| KNN | 0.58 | 0.66 | 58.25 |
| SVM | 0.53 | 0.56 | 53.39 |
| Vote | 0.61 | 0.68 | 61.16 |

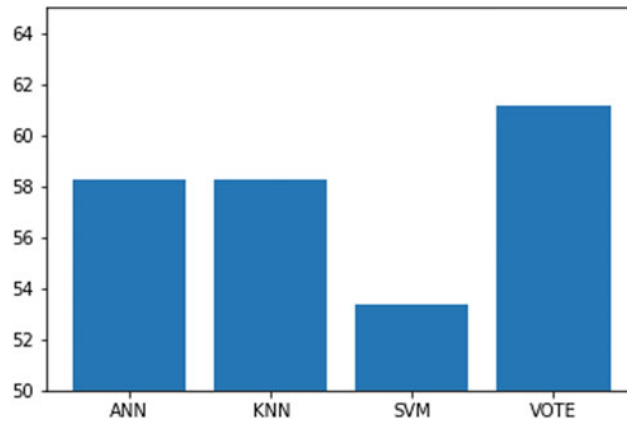**Fig. 3** Multi class classification results by percentage split

**Table 3** Confusion matrix obtained by percentage split

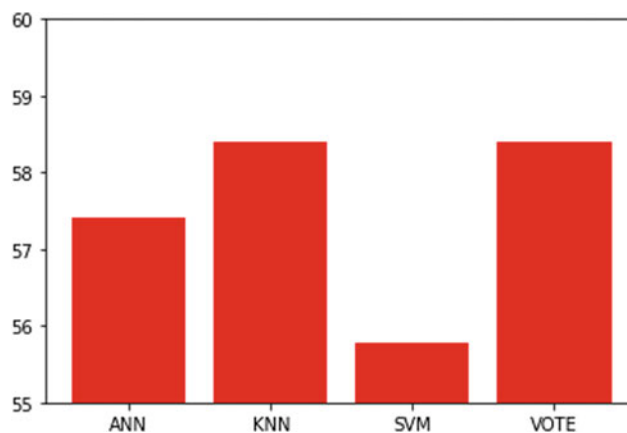| | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| 0 | 48 | 2 | 1 | 0 | 0 |
| 1 | 9 | 5 | 2 | 0 | 0 |
| 2 | 1 | 4 | 8 | 0 | 0 |
| 3 | 1 | 12 | 5 | 2 | 0 |
| 4 | 0 | 0 | 3 | 0 | 0 |

**Fig. 4** Multi class classification results by 10-fold cross validation

**Table 4** Multi class classification results by 10-fold cross validation

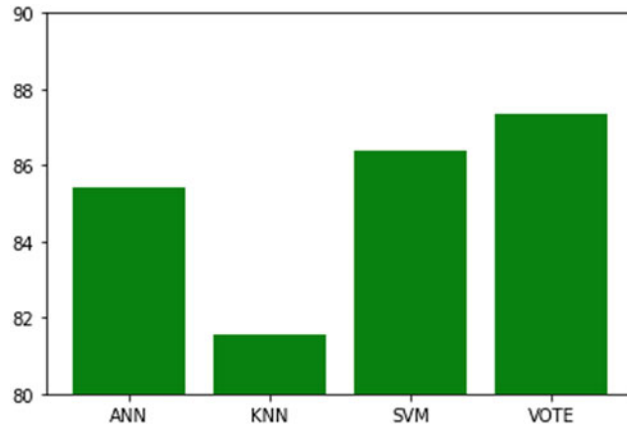| Algorithm | Sensitivity | Precision | Accuracy |
|---|---|---|---|
| ANN | 0.57 | 0.52 | 57.42 |
| KNN | 0.58 | 0.52 | 58.41 |
| SVM | 0.55 | 0.48 | 55.77 |
| Vote | 0.58 | 0.52 | 58.41 |



**Fig. 5** Binary classification results by percentage split

**Table 5** Confusion matrix obtained by 10-fold cross validation

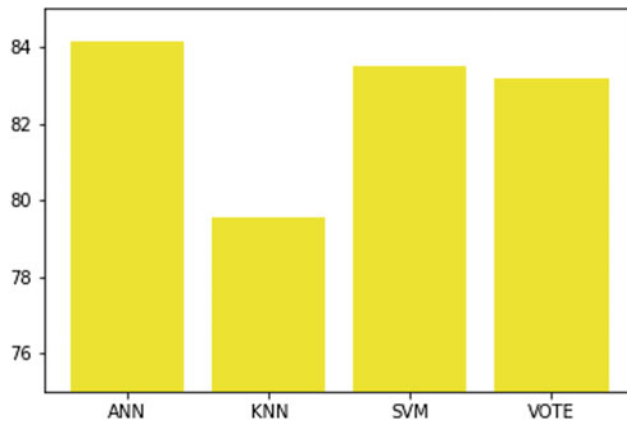| | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| 0 | 150 | 8 | 4 | 2 | 0 |
| 1 | 28 | 9 | 13 | 5 | 0 |
| 2 | 9 | 7 | 12 | 8 | 0 |
| 3 | 7 | 7 | 15 | 6 | 0 |
| 4 | 1 | 2 | 8 | 2 | 0 |



**Fig. 6** Binary classification results by 10-fold cross validation

**Table 6** Binary classification results by percentage split

| Algorithm | Sensitivity | Precision | Accuracy |
|---|---|---|---|
| ANN | 0.85 | 0.85 | 85.43 |
| KNN | 0.81 | 0.81 | 81.55 |
| SVM | 0.86 | 0.86 | 86.40 |
| Vote | 0.87 | 0.87 | 87.37 |

**Table 7** Binary classification results by 10-fold cross validation

| Algorithm | Sensitivity | Precision | Accuracy |
|---|---|---|---|
| ANN | 0.84 | 0.84 | 84.15 |
| KNN | 0.79 | 0.79 | 79.53 |
| SVM | 0.83 | 0.83 | 83.49 |
| Vote | 0.83 | 0.83 | 83.16 |

Table 5 and Fig. 6 show results obtained by binary classification and 10-fold cross validation. Unlike all previous cases, here majority voting did not achieve highest accuracy. Highest accuracy 84.15 is obtained by ANN. As with percentage split, KNN achieved lowest accuracy (Tables 6 and 7).

For binary classification, all results obtained by percentage split are higher than those obtained by cross validation.

## 5   Conclusion

In this paper we presented application of artificial neural network, k nearest neighbor and support vector machine on dataset with 14 attributes and 303 instances.

We evaluated difference between results obtained by algorithms applied separately and majority voting as ensemble learning. Problem is solved in two ways: as multiclass and binary classification, using two types of evaluation: percentage split and cross validation. We applied 66:34 percentage split and 10-fold cross validation.

In three of the four cases majority voting had highest accuracy equal to 61.15 for multiclass classification and 87.37 for binary classification. Only in binary classification with 10-fold cross validation single algorithm, ANN, outperformed ensemble learning method. This superiority of the ensemble learning over single algorithms could be explained by the fact that ensemble learning combines the best of all algorithms and gives single result. Also, results for binary classification are higher than those for multiclass classification. As the reason we could conclude number of classes that are available in decision making, since it is harder to 'learn' with five than with two outputs.

When it comes to data split, generally higher values of specificity, precision and accuracy are obtained by percentage split than by 10-fold cross validation.

Therefore, we may conclude that majority voting outperforms algorithms used separately and percentage split gives better results compared to 10-fold cross validation.

**Conflict of Interest Declaration** The authors declare that they have no conflict of interest.

## References

1. Luo, G.: MLBCD: a machine learning tool for big clinical data. Health Inform. Sci. Syst. **3** (2015)
2. Belle, A., Thiagarajan, R., Reza Soroushmehr, S.M., Navidi, F., Beard, D.A., Najarian, K.: Big data analytics in healthcare. Biomed Res. Int. **2015** (2015)
3. Corbett, E.: Real-world benefits of machine learning in healthcare. Health Catalyst. [Online]. Available: https://www.healthcatalyst.com/clinical-applications-of-machine-learning-in-healthcare (25 Apr 2017). Accessed 12 Mar 2018
4. Krieger, L.M.: Google computers trained to detect cancer. The Mercury News. [Online]. Available: https://www.mercurynews.com/2017/03/03/googlecomputers-trained-to-detect-cancer/ (03 Mar 2017). Accessed 13 Jan 2019
5. Stanford University: Artificial intelligence used to identify skin cancer | Stanford News. Stanford News. [Online]. Available: https://news.stanford.edu/2017/01/25/artificial-intelligence-used-identify-skin-cancer/ (25 Jan 2017). Accessed 13 Jan 2019
6. Wang, P., et al.: Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy. Nat. Biomed. Eng. **2**(10), 741–748 (2018)
7. WHO | World Health Organization (Mar 2018)
8. WHO | Cardiovascular Diseases (CVDs) (May 2017)
9. American Heart Association: Building healthier lives, free of cardiovascular diseases and stroke. [Online]. Available: http://www.heart.org/HEARTORG/. Accessed 23 Dec 2017
10. Kim, G.B.: Psychosocial adjustment and quality of life of adolescents and adults with congenital heart disease. Korean J. Pediatr. **57**(6), 257–263 (2014)
11. Ensemble Methods: Elegant Techniques to Produce Improved Machine Learning Results. Toptal Engineering Blog. [Online]. Available: https://www.toptal.com/machine-learning/ensemble-methods-machine-learning. Accessed 06 Mar 2018

12. Singh, G., Bagwe, K., Shanbhag, S., Singh, S., Devi, S.: Heart disease prediction using Naïve Bayes. Int. Res. J. Eng. Technol. (IRJET) **03**(04) (Mar 2017)

13. Devi, S.K., Krishnapriya, S., Kalita, D.: Prediction of heart disease using data mining techniques. Indian J. Sci. Technol. **9**(39) (2016)

14. El-Bialy, R., Salamay, M.A., Karam, O.H., Khalifa, M.E.: Feature analysis of coronary artery heart disease data sets. Procedia Comput. Sci. **65**, 459–468 (2015)

15. Venkatalakshmi, B., Shivsankar, M.V.: Heart disease diagnosis using predictive data mining. Int. J. Innov. Res. Sci. Eng. Technol. (2014)

16. Jabbar, M.A., Deekshatulu, B.L., Chandra, P.: Classification of heart disease using K-nearest neighbor and genetic algorithm. Procedia Technol. **10**, 85–94 (2013)

17. UCI Machine Learning Repository: Heart Disease Data Set. [Online]. Available: http://archive.ics.uci.edu/ml/datasets/heart+Disease. Accessed 16 Dec 2017

18. Brownlee, J.: An introduction to feature selection. Machine Learning Mastery. [Online]. Available: https://machinelearningmastery.com/an-introductionto-feature-selection/ (06 Oct 2014). Accessed 02 Sep 2018

19. Precision and Recall—Wikipedia. [Online]. Available: https://en.wikipedia.org/wiki/Precision_and_recall. Accessed 02 Sep 2018