

Hybrid Ensemble Framework for Heart Disease Detection and Prediction

Elham Nikookar

Department of Computer Engineering, Faculty of
Engineering
Shahid Chamran University of Ahvaz
Ahvaz, Iran

Ebrahim Naderi

Computer Department,
University of Applied Science and Technology,
Ahvaz, Iran

Abstract—Data mining techniques have been widely used in clinical decision support systems for detection and prediction of various diseases. As heart disease is the leading cause of death for both men and women, detection and prediction of the heart disease is one of the most important issues in medical domain and many researchers developed intelligent medical decision support systems to improve the ability of the CAD systems in diagnosing heart disease. However, there are almost no studies investigating capabilities of hybrid ensemble methods in building a detection and prediction model for heart disease. In this work, we investigate the use of hybrid ensemble model in which a more reliable ensemble than basic ensemble models is proposed and leads to better performance than other heart disease prediction models. To evaluate the performance of proposed model, a dataset containing 278 samples from SPECT heart disease database is used that after applying the model on the data, 96% of classification accuracy, 80% of sensitivity and 93% of specificity are obtained that indicates acceptable performance of the proposed hybrid ensemble model in comparison with basic ensemble model as well as other state of the art models.

Keywords—Data mining; hybrid ensemble; base classifier; classification accuracy; sensitivity; specificity

I. INTRODUCTION

The World Health Organization has estimated that 12 million deaths occur worldwide, every year due to the heart diseases [1]. Although, in the last few decades many computational tools have been designed to improve the abilities of physicians for making decisions about condition of disease in their patients [2], low performance of current heart disease detection models is remained a matter of concern and potential of data mining algorithms which are motivated by the need of an expert system, have not be highlighted in any research yet.

Artificial intelligence techniques as a subfield of data mining have been increasingly used in solving problems in medical domains such as in oncology, urology, liver pathology, cardiology, gynecology, thyroid disorders and perinatology [2]. The primary concern of artificial intelligence in medicine is construction of an intelligent system that can assist a medical doctor in performing expert diagnosis as well as predicting probability of a disease in a patient more accurately. Besides, artificial intelligence algorithms have great potential for exploring the hidden patterns in the datasets of the various disease related subjects by adjusting the data mining model for utilizing such patterns for clinical diagnosis

[1] and this potential has led to building expert systems that can be used in CAD systems for prediction and detection of diseases in patients. One of the concepts that have been emerged in recent years is the idea of combining classifiers as a new direction for the improvement of the performance of individual classifiers [3]. These classifiers could be based on a variety of classification methodologies and could achieve different rate of correctly classified samples. Such classifiers which are called ensemble classifiers have potential to lead to an increase in generalization performance by combining several base or weak classifiers and train them on the same task [4]. However, although in recent years, better models of ensemble classifiers such as hybrid ensemble classifier which have been proved to achieve better performance than basic ensemble algorithms has been introduced [5], there are almost no studies investigating application of hybrid ensemble models and their feasibilities in heart disease domain. Thus, in this study, we evaluate the performance of a hybrid ensemble model which uses five popular classification methods including Naïve Bayes, k-NN, Random Tree, SVM and Bayes Net as base classifiers and takes benefits of aggregating all these classifiers by forwarding their results to a novel fuser classifier which is chosen in this study between Adaboost, LogitBoost, MLP and Random Forest for the diagnosis of the heart disease disorders. To evaluate the performance of the proposed model, a comparative study is realized by using a dataset containing 267 samples which is available in public UCI Repertory website [6]. We finally show that the proposed method is capable of being used as a more powerful tool to assist the medical doctor in detection and prediction of the heart disease than the basic ensemble models as well as other state of the art models.

This paper is organized as follows. Section II presents the dataset that is used to train, test and evaluate the proposed model. In Section III a number of previous studies in heart disease detection and prediction domain is discussed which culminates with an identification of the knowledge gap and inconsistencies in the literature. Section IV explicitly explains the proposed model and Section V provides the performance evaluation measures used in this study. In Section VI single base classifier model which is investigated to be compared with proposed model is introduced and in Section VII experimental results are provided. Section VIII presents a general discussion of the study. Section IX concludes the study and Section X provides the recommendations for future studies.

II. DATASET

SPECT heart disease dataset is used in this paper which is available on university of California, Irvine (UCI) machine learning dataset repository [6]. The dataset is provided for investigating diagnose of cardiac Single Proton Emission Computed Tomography (SPECT) images using machine learning algorithms. SPECT, or less commonly, SPET, is a nuclear medicine tomographic imaging technique using gamma rays. It is very similar to conventional nuclear medicine planar imaging using a gamma camera (that is, scintigraphy). However, it is able to provide true 3D information. This information is typically presented as cross-sectional slices through the patient, but can be freely reformatted or manipulated as required.

SPECT heart disease dataset was obtained from Medical College of Ohio, OH, U.S.A. The database of 267 SPECT image sets (patients) was processed to extract features that summarize the original SPECT images. As a result, 22 continuous feature patterns were created for each patient. All continuous attributes have integer values from the 0 to 100 but were further processes to obtain 22 binary feature patterns. Each of the patients is classified into one of two categories: normal and abnormal. SPECT dataset was firstly utilized by Kurgan et al. [7] where they used CLIP3 algorithm which used to generate classification rules from Features. The performance evaluation of their proposed model was evaluated by classification accuracy and the maximum value that they achieved was 90.4% in their study.

III. BACKGROUND

Classification algorithms are generally very useful for medicinal issues, especially when applied for the heart disease detection and prediction purposes [8]-[16]. Many machine learning algorithms are applied in the medical domain in the course of recent decades. A large portion of these applications are specific and include machine learning procedures like using data mining for identification and detection of disease in patients [7] and application of neural network rules for the prediction of breast cancer [17]. For example, in [18] an intelligent model is proposed for the detection of heart disease based on wavelet packet neural networks (WPNN) and they reported 94% of correct classification rate for abnormal and normal subjects. In [11] a system is proposed for diagnosis and prediction of heart disease based on Genetic Neural Network Using Risk Factors. In [9] the use of least-square support vector machines (LS-SVM) classifier for improving the performance of the proposed model of [13] is investigated. However, according to what previous studies reported, they did not investigate the use of hybrid ensemble methods to predict the occurrence of heart disease based on SPECT images of patients. Lack of research studies on this topic makes it unclear whether the hybrid ensemble models are capable of providing a model that utilizes the power of ensemble model by merging initial features of patients and predicted class labels by base classifiers. Therefore, the present study is focused on the idea of hybrid ensemble models and investigates the effectiveness of such models on the performance of a heart disease detection and prediction system.

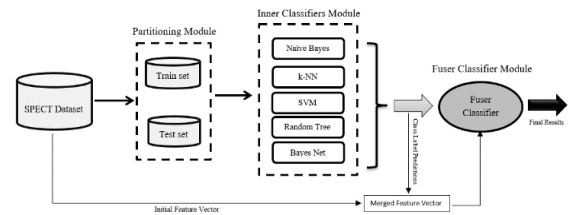


Fig. 1. The general flowchart of proposed hybrid ensemble model.

IV. METHOD

The aim of this paper is to propose a hybrid ensemble model for heart disease detection and prediction which focuses on predicting labels of each SPECT image based on feature vector of the images and the labels that base classifiers assign to each image. To facilitate understanding of the proposed framework, in this section we describe the details of layout of the proposed model. A schematic illustration of proposed hybrid ensemble model can be seen in Fig. 1. It consists of three modules, including *partitioning module*, *inner classifiers module* and *fuser module*. The initial dataset is first given to *partitioning module* to produce train and test subsets and prepare them for the next module. In *inner classifiers module* different classification algorithms are applied on the train and test datasets to produce input data for *fuser module* in which results of base classifiers next to initial feature vector of samples are considered simultaneously for building and adjusting components of the final classifier. In the rest of this section, a brief description of each component is given.

A. Partitioning Module

This module divides the initial dataset into test and train subsets by assigning 80 samples to train set and 187 samples to test set and provides mutually exclusive datasets which share no instance with each other and provides initial data for base classifiers in the next module.

B. Inner Classifiers Module

This module is constructed using five classification algorithms as base or weak classifiers including Naïve Bayes, k-NN, Random Tree, SVM and Bayes Net. All these base classifiers are applied on the train data using 10-fold cross validation as model validation technique to be adjusted for the best possible prediction about healthy or unhealthy situation of a patient. The reason of considering odd number of classifiers in *inner classifiers module* is based on the pigeonhole principle [19], which states that for natural numbers k and m , if $n=km+1$ objects are distributed among m sets, at least one of the sets will contain at least $k+1$ objects. For arbitrary n and m , it generalizes to $k+1=\lfloor(n-1)/m\rfloor+1$, where $\lfloor \cdot \rfloor$ is the floor function. It means that in the two-class problem (healthy 0, unhealthy 1) in which each classifier has to give its vote for the class of a sample, there is a need to have an odd number of classifiers to avoid equal 0 and 1 predictions for a sample. This odd number is considered five in this study. The increasing number of classifiers may obviously result in finding a more powerful model for the data but it has the risk of overfitting the model on the specific data which is used in this study. After applying all runs of 10-fold cross validation, test dataset is given to *inner classifiers module* to assign each test sample five labels by five base classifiers. To provide

input data for fuser classifier in next module, these labels are added to feature vector of test samples which leads to generating a new feature vector for each test sample including 22 binary features from initial feature vector and 5 features from *inner classifiers module* of proposed model.

C. Fuser Module

After training and testing the five classifiers in *inner classifiers module*, a new feature vector is built with 27 features including 5 predicted class labels by five base classifiers plus 22 initial features of samples. Then, the new dataset is used to find an optimal fuser classifier for the model. The candidates for fuser classifier are Adaboost, LogitBoost, MLP and Random Forest. As the fuser classifier needs to be trained to fit the data in the best form, the test dataset produced in *partitioning module* is divided itself into test and train subsets using a stratified training-test partition (80-20) and 10-fold cross validation is used as model selection technique in *fuser module* to adjust the fuser classifier and complete the hybrid ensemble model. The final result of the model is then produced for all test samples.

V. PERFORMANCE EVALUATION MEASURES

Performance evaluation is mandatory in all automated disease recognition systems and is conducted in this study to evaluate the ability of base classifiers as well as proposed hybrid ensemble model for predicting possibility of heart disease in patients based on SPECT images. Although precision and recall are more common in general data mining tasks, in medical domain, researchers prefer to assess how much sensitive and specific their proposed model is and the standard evaluation measures are sensitivity and specificity. Actually, in clinical context, a more sensitive model is preferable as the cost of overlooking a positive sample is very high and a more specific model is preferable as the cost of registering a sample as positive for the samples that are not the target of testing is very high [20].

$$Sensitivity = tp / (tp + fn)$$

$$Specificity = tn / (tn + fp)$$

The classification accuracy is also considered as evaluation measure in this study as it facilitates comparison of the results of present study with other state of the art models. The classification accuracy, CA, depends on the number of samples correctly classified (true positives plus true negatives) and is evaluated by the formula:

$$CA = (t/n).100$$

where t is the number of sample cases correctly classified, and n is the total number of sample cases.

VI. SINGLE BASE CLASSIFIER MODEL

To compare the results of our proposed model with the situation in which only a single base classifier is used, such as only SVM is investigated, this study separately applied all the five classifiers on the dataset. General flowchart of applying single base classifier is illustrated in Fig. 2. Same dataset diving procedure like *partitioning module*, i.e. 30% for train data and 70% for test data, as well as 10-fold cross validation have been used for providing data for each single base classifier training and testing.

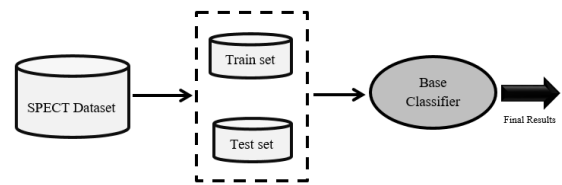


Fig. 2. The general flowchart of applying a single base classifier on the dataset.

VII. EXPERIMENTAL RESULTS

Weka, which is a collection of machine learning algorithms for data mining tasks [21] is used to train, test and evaluate the proposed model as it has two important characteristics; it is a free software system and it uses ARFF files that can be easily used and modified without data format problems. The results of applying base classifiers as well as results of experiments conducted to choose the best fuser classifier will be discussed in part 1 of this section. In part 2, results of applying different classifiers as fuser classifier will be investigated and a comparison between single base classifier model introduced in Section VI and the proposed hybrid ensemble model will be discussed in part 3. Comparing results of applying basic ensemble with results of the proposed model is conducted in part 4. In addition, a comparison between the proposed hybrid ensemble model and other heart disease detection and prediction systems will be discussed in part 5.

1) Results of Applying Single base Classifiers

As we mentioned earlier, five well-known classification algorithms including Naïve Bayes, k-NN, SVM, Random Tree and Bayes Net were used in *inner classifiers module* for constructing the core of proposed hybrid ensemble model. The experimental results of applying each base classifier on the train data is given in Table I.

As shown in Table I, the results indicate that the best base classifiers are k-NN and Random Tree considering sum value of three evaluation measures as decision criterion, However, we do not ignore predicted labels by any of base classifiers and in the next step, vote of each base classifier which is a predicted class label is kept to be used in *fuser module* for the purpose of constructing new feature vector for the hybrid ensemble model.

TABLE I. RESULTS OF APPLYING BASE CLASSIFIERS ON THE TRAIN DATA.

Method	Sensitivity	Specificity	CA
Naïve Bayes	55%	75%	69%
k-NN	63%	75%	76%
SVM	63%	48%	75%
Random Tree	68%	62%	78%
Bayes Net	55%	75%	69%

TABLE II. RESULTS OF APPLYING DIFFERENT FUSER CLASSIFIERS ON THE DATA.

Method	Sensitivity	Specificity	CA
Adaboost	71%	93%	85%
LogitBoost	85%	77%	93%
MLP	80%	93%	96%
Random Forest	77%	77%	90%

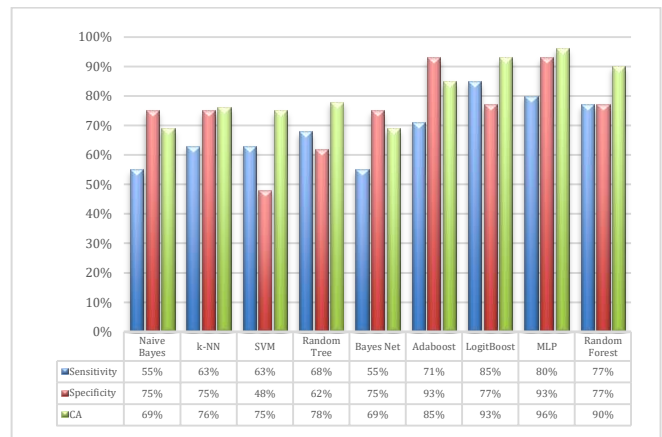
2) *Choosing best Fuser Classifier*

For selection of best fuser classifier many choices were available among diverse collection of classification algorithms. Among all choices, Adaboost, LogitBoost, MLP and Random Forest were chosen as they proved to produce acceptable results in most of the machine learning models. The experimental results of applying different fuser classifiers can be observed in Table II. It is needed to point that for each fuser classifier different configurations has been tested and the best result of each classifier is inserted in Table II.

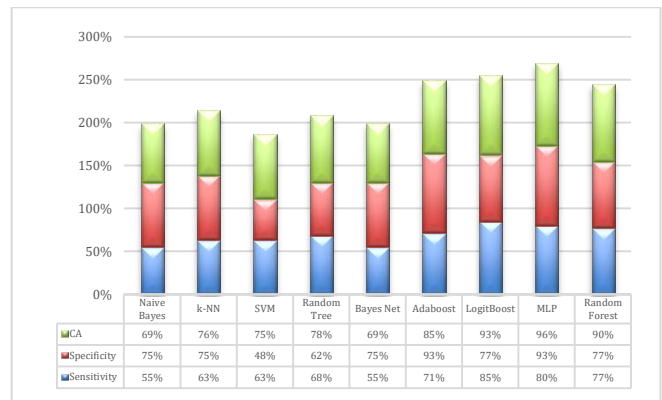
Based on the results, MLP is the best performing candidate to be chosen as fuser classifier. The parameters of the MLP that we applied on the data were set as follows: The backpropagation learning algorithm has been used in the feedforward single hidden layer of the neural network. The algorithm used for training the proposed MLP is the Levenberg–Marquardt (LM) algorithm [22]. A tangent sigmoid transfer function has been used for both the hidden layer and the output layer of the model. Besides, we used 10 neurons in the hidden layer, the initial weights were chosen randomly and in regression node, logistic regression was used.

3) *Compare Hybrid Ensemble Model with base Classifiers*

In Fig. 3, the results of applying different base classifiers as well as results after applying fuser classifier methods in *fuser module* can be observed. It is clear that the hybrid ensemble model enhances results of base classifiers on the data and there is a considerable difference between results of best performing base classifiers which are k-NN and Random Tree and the results of best fuser classifier, i.e. MLP. Therefore, the proposed hybrid ensemble model has its strength from both powerful base classifiers in *inner classifiers module* and fuser classifier which incorporates initial features of samples with predictions of base classifiers for samples. In fact, the idea of applying fuser classifiers for building up an effective ensemble classifier, in line with the idea of adding predictions of base classifiers to initial feature vector of samples, leads to final results of the proposed model. From the experimental results that are given in Fig. 3 in two charts, we conclude that the proposed hybrid ensemble model outperforms all base classifiers in terms of sensitivity, specificity and classification accuracy.



(a)



(b)

Fig. 3. Comparison between results of fuser classifiers with single base classifiers. Chart (a) shows each evaluation measure with separate bar and chart (b) shows all of the measures with one bar.

4) *Compare Hybrid Ensemble Model with basic Ensemble Model*

A basic ensemble in this study means an ensemble similar to proposed hybrid ensemble model with the difference that in basic ensemble model, feature vector of the samples which are fed to final module of the model only includes class labels predicted by base classifiers in *inner classifiers module* and does not include initial features of samples. The results of applying basic ensembles on the test data is shown in Table III. The results indicate that the idea of merging features with predicted class labels led to construction a model with better performance.

TABLE III. COMPARISON BETWEEN RESULTS OF APPLYING HYBRID ENSEMBLE (HE) AND BASIC ENSEMBLE (BE) ON THE TEST DATA

Method	Sensitivity		Specificity		CA	
	HE	BE	HE	BE	HE	BE
Adaboost	71%	53%	93%	80%	85%	69%
LogitBoost	85%	62%	77%	70%	93%	60%
MLP	80%	68%	93%	74%	96%	68%
Random Forest	77%	61%	77%	70%	90%	72%

TABLE IV. COMPARISON BETWEEN PROPOSED MODEL AND OTHER HEART DISEASE DETECTION AND PREDICTION MODELS.

Author	Method	CA
[23]	Naïve Bayes	81%
[10]	Fuzzy-AIRS-Knn based System	87%
[13]	Neural Network Ensemble	89%
[7]	CLIP3	90%
[18]	wavelet packet neural networks (WPNN)	94%
Proposed Model	Hybrid Ensemble Model	96%

5) Comparison with other Heart Disease Detection Methods

Although the experiment has achieved acceptable results by building a hybrid ensemble model, another important challenge is to compare current study with other previous methods. Related studies reporting same evaluation measures to the present study has been searched. The majority of the previous studies applied their models on private datasets and reported the results in different forms as there is no standard for this process. With all this among similar studies, as shown in Table IV, the proposed approach has provided better performance than the other techniques regarding to the classification accuracy which is the general performance measure that is used in all related studies.

VIII. DISCUSSION

The ability of an artificial intelligence model in predicting the possibility of heart disease is imperative for decreasing the mortality rate of heart disease. The ability in this study is expressed in terms of evaluation measures including sensitivity, specificity and classification accuracy that in our best configuration, the experimental results respectively show the values of 80%, 93% and 96% for these evaluation measures. This study highlights two important aspects. First, the effectiveness of using an ensemble classifier instead of base classifiers may be obvious. Second and the more important, the effectiveness of considering a combination of initial features of samples and class labels of samples predicted by base classifiers as the feature vector of fuser classifier instead of only considering predicted class labels by base classifiers which is common in basic ensemble classifiers. In the other words, the second aspect considers effectiveness of using hybrid ensemble classifier instead on basic ensemble classifier. For the first aspect, the use of hybrid ensemble classifier for heart disease detection and prediction has reached to 80%, 93% and 96% for sensitivity, specificity and classification accuracy which is 12%, 18% and 18% more than results of best base classifier (assuming highest values of measures for k-NN and Random Tree base classifiers). For the second aspect, based on Table III, it can be seen that better performance is achieved by applying a hybrid ensemble classifier instead of a basic ensemble classifier and the results show 12%, 19% and 18% improvement is sensitivity, specificity and classification accuracy respectively. These results show that the idea of proposed hybrid ensemble model has improved the ability and effectiveness of heart disease detection and prediction artificial intelligence models.

IX. CONCLUSIONS

The proposed heart disease detection and prediction model enables the physician to predict and diagnose the heart disease by investigating and analyzing Single Proton Emission Computed Tomography (SPECT) images of patients. The artificial intelligence models that use SPECT images have been underscored in the previous studies. However, there is a limited number of works that underscore use of a hybrid ensemble classifier in a heart disease detection and prediction artificial intelligence model. Therefore, this study introduces a new approach that merges initial features of samples and base classifier predictions to produces a new feature vector for fuser classifier. It culminates with the formulation of a new model, which is considered as a novel of the present study.

In order to build a reliable model, this study investigated different fuser classifiers and considered comparison between basic ensemble and hybrid ensemble as well as comparison between hybrid ensemble and base classifiers. The results obtained from different configurations of the model indicate that the proposed model is a more reliable system that can support clinical decision makers by providing more reliable information. The proposed model is an effective artificial intelligence model for predicting heart disease, especially in terms of sensitivity and specificity that are clinically important evaluation measures. This improvement would increase the performance of the heart disease CAD systems in the clinical environments. As a conclusion, this study confirms that merging initial features of samples with predicted class labels of samples by different classification algorithms would be advantageous for the clinical decision makers.

X. FUTURE WORKS

Our study raises a number of opportunities for future researches on heart disease prediction models. As mentioned in Section I, this study uses five classifiers in inner classifiers module. This limitation is due to a tradeoff between model simplicity and maximum possible values of evaluation measures. Although this study outlines the model simplicity, however, it is a challenge to add more classifiers to reach better performance. Future researches may also tackle the proposed model by applying more fuser classifiers. In addition, another opportunity for future researches would be extending the proposed model for other types of diseases.

REFERENCES

- [1] Soni, J., et al., Predictive data mining for medical diagnosis: An overview of heart disease prediction. *International Journal of Computer Applications*, 2011. 17(8): p. 43-48.
- [2] Sarwar, A., V. Sharma, and R. Gupta, Hybrid ensemble learning technique for screening of cervical cancer using Papanicolaou smear image analysis. *Personalized Medicine Universe*, 2015. 4: p. 54-62.
- [3] Mohapatra, S., D. Patra, and S. Satpathy, An ensemble classifier system for early diagnosis of acute lymphoblastic leukemia in blood microscopic images. *Neural Computing and Applications*, 2014. 24(7-8): p. 1887-1904.
- [4] A.R., E. Naderi, and A. Osareh. Parallel weak learners, a novel ensemble method. in *Computational Intelligence and Computing Research (ICCIC)*, 2010 IEEE International Conference on. 2010. IEEE.
- [5] Woźniak, M., M. Graña, and E. Corchado, A survey of multiple classifier systems as hybrid systems. *Information Fusion*, 2014. 16: p. 3-17.

- [6] Dua, D. and E.K. Taniskidou, UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. 2017.
- [7] Kurgan, L. and K.J. Cios. Ensemble of classifiers to improve accuracy of the CLIP4 machine-learning algorithm. in *Sensor Fusion: Architectures, Algorithms, and Applications VI*. 2002. International Society for Optics and Photonics.
- [8] Thanigaivel, R. and K.R. Kumar, Review on heart disease prediction system using data mining techniques. *Asian Journal of Computer Science and Technology (AJCST)* Vol, 2015. 3: p. 68-74.
- [9] Çomak, E., A. Arslan, and İ. Türkoğlu, A decision support system based on support vector machines for diagnosis of the heart valve diseases. *Computers in Biology and Medicine*, 2007. 37(1): p. 21-27.
- [10] Polat, K., et al. A new classification method to diagnosis heart disease: supervised artificial immune system (AIRS). in *proceedings of the turkish symposium on artificial intelligence and neural networks (TAINN)*. 2005.
- [11] Amin, S.U., K. Agarwal, and R. Beg. Genetic neural network based data mining in prediction of heart disease using risk factors. in *Information and Communication Technologies (ICT), 2013 IEEE Conference on*. 2013. IEEE.
- [12] Das, R. and A. Sengur, Evaluation of ensemble methods for diagnosing of valvular heart disease. *Expert Systems with Applications*, 2010. 37(7): p. 5110-5115.
- [13] Das, R., I. Turkoglu, and A. Sengur, Effective diagnosis of heart disease through neural networks ensembles. *Expert systems with applications*, 2009. 36(4): p. 7675-7680.
- [14] Dai, W., et al., Prediction of hospitalization due to heart diseases by supervised learning methods. *International journal of medical informatics*, 2015. 84(3): p. 189-197.
- [15] Nahar, J., et al., Computational intelligence for heart disease diagnosis: A medical knowledge driven approach. *Expert Systems with Applications*, 2013. 40(1): p. 96-104.
- [16] Kaur, B. and W. Singh, Review on heart disease prediction system using data mining techniques. *International journal on recent and innovation trends in computing and communication*, 2014. 2(10): p. 3003-3008.
- [17] Kurgan, L.A., et al., Knowledge discovery approach to automated cardiac SPECT diagnosis. *Artificial intelligence in medicine*, 2001. 23(2): p. 149-169.
- [18] Turkoglu, I., A. Arslan, and E. Ilkay, An intelligent system for diagnosis of the heart valve diseases with wavelet packet neural networks. *Computers in Biology and Medicine*, 2003. 33(4): p. 319-331.
- [19] Trybulec, W.A., Pigeon hole principle. *Formalized Mathematics*, 1990. 1(3): p. 575-579.
- [20] Altman, D.G. and J.M. Bland, Diagnostic tests. 1: Sensitivity and specificity. *BMJ: British Medical Journal*, 1994. 308(6943): p. 1552.
- [21] Hall, M., et al., The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter*, 2009. 11(1): p. 10-18.
- [22] Das, R., A comparison of multiple classification methods for diagnosis of Parkinson disease. *Expert Systems with Applications*, 2010. 37(2): p. 1568-1572.
- [23] Cheung, N., Machine learning techniques for medical analysis. School of Information Technology and Electrical Engineering. 2001, B. Sc. Thesis, University of Queensland.