

Impact of K-Means and DBSCAN Clustering on Supervised Learning for Heart Disease Prediction

¹Pulugu Dileep, ²Kunjam Nageswara Rao, ³Prajna Bodapati, ⁴Sitaratnam Gokuruboyina, ⁵Revathy Peddi

¹Associate Professor, Department of CSE, Malla Reddy College of Engineering and Technology, Telangana

²Professor, Department of CS & SE, AUCE(A) at Andhra University, Visakhapatnam, Andhra Pradesh

³Professor, Department of CS & SE, AUCE(A) at Andhra University, Visakhapatnam, Andhra Pradesh

⁴Professor, Department of CSE, LENDI Institute of Engineering and Technology, Vizianagaram, Andhra Pradesh

⁵Assistant Professor Department of CSE, ACE Engineering College, Hyderabad, Telangana, India.

¹myresearchwork19@gmail.com, ²kunjamnag@gmail.com, ³prajna.mail@gmail.com,

⁴sitagokuruboyina@gmail.com, ⁵revathy5813@gmail.com

Article History: Received: 10 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 20 April 2021

Abstract: Cardiovascular diseases (CVDs) are the main cause of death of around 17.9 million across the globe every year. Different heart ailments or conditions exist that lead to death. Early detection of heart disease can help in preventing death rate. Data driven approaches with Artificial Intelligence (AI) innovations are being used for Clinical Decision Support System (CDSS). From the literature, it is found that most of the algorithms are based on supervised learning that need training set for learning process. Often supervised learning have revealed limitations as they wholly depend on the quality of training data. Feature selection has been around for leveraging supervised machine learning techniques. However, there is inadequate research on the unsupervised learning methods such as clustering. Nevertheless, they around round to group objects with similarities and thus have inherent knowledge that can add value to features extracted. Provided this fact, in this research, we proposed a framework known as Hybrid Machine Learning for Heart Disease Prediction (HML-HDP). The framework has provision for unsupervised learning followed by supervised learning. In other words, the unsupervised learning could lead to features that complement feature selection method for improving accuracy of heart disease prediction. A prototype application is built in order to evaluate the framework. The empirical results are compared with the methods in state of the art. SVM with DBSCAN showed Highest performance with precision 0.96, recall 0.93 and F-Measure 0.9447. The results revealed that the proposed hybrid framework performs better than existing methods.

keywords –Heart disease prediction, machine learning, artificial intelligence, CDSS

1. Introduction

Heart disease prediction is an important research that has been around for number of years. However, with technological innovations and emergence of data science, there are ever possibilities to improve the state of the art in prediction of heart diseases. Thus it is essential to detect heart diseases early in order to reduce mortality rate. Many researchers contributed towards this end. Associative classification is explored in [2] for better prediction of the disease. Supervised bi-clustering is used in [3] for exploiting benefits of unsupervised learning. Particle Swarm Optimization (PSO) and ensemble learning methods are combined for improved prediction accuracy in [5]. The concept of computational modelling is investigated in [6] for prediction of disease.

The literature is rich in modern innovations as well. In [7] Internet of Things (IoT) technology is integrated with the healthcare system for diagnosis of heart disease. Deep learning concept is studied and used for heart disease diagnosis in [20] and [23]. With innovations such as cloud computing, big data analytics is employed in [14], [21] and [25]. Fuzzy based decision support system for clinical diagnosis of heart diseases is the main focus in [8], [12] and [18]. From the literature, it is found that there are unsupervised and supervised learning methods widely used for predicting heart diseases. However, there is need for having a framework that combines and reaps benefits of both supervised and unsupervised learning methods. This gap is considered in this paper. Our contributions in this paper are as follows.

1. Proposed a heart disease prediction framework known as Hybrid Machine Learning for Heart Disease Prediction (HML-HDP). The framework integrates both supervised and unsupervised prediction models.
2. K-Means and DBSCAN are the two clustering algorithms combined with many supervised prediction models such as Naïve Bayes (NB), Random Forest (RF) and Support Vector Machine (SVM).
3. A prototype application is built using Python data science platform for realizing the proposed framework.

The remainder of the paper is structured as follows. Section 2 reviews literature on both supervised and unsupervised prediction models. Section 3 presents the proposed framework for prediction of heart diseases.

Section 4 presents evaluation methodology. Section 5 gives details of experimental setup. Section 6 provides experimental results while section 7 concludes the paper and gives directions for possible future scope of the research.

2. Related work

This section provides literature findings on usage of clustering and classification techniques for heart disease prediction. Aljawarneh *et al.* [1] proposed a methodology for predicting disease using gene profile classification. It is a supervised learning technique. Singh *et al.* [2] on the other hand explored associative classification for efficient prediction of heart diseases. Nezhad *et al.* [3] focused on the problem of precision medicine. They used bi-clustering approach in order to achieve better results than state of the art. Antiochos *et al.* [4] used a novel approach for heart disease diagnosis. They combined both genetic risk scores and parental history to achieve this. Yekkala *et al.* [5] used Particle Swarm Optimization (PSO) technique as wrapper and also explored ensemble learning for heart disease prediction. They found that ensemble approach is better than individual prediction models.

Mittal *et al.* [6] studied computational modelling to ascertain cardiac hemodynamics. Kumar *et al.* [7] proposed an architecture using Internet of Things (IoT) for heart disease prediction in distributed environment. Paul *et al.* [8] used Genetic Algorithm (GA) for fuzzy decision support to diagnose heart disease. Babic *et al.* [9] employed both predictive and descriptive analysis to diagnose heart diseases. Verma and Sood [10] proposed a solution for heart disease prediction using IoT and cloud based system. Malav *et al.* [11] used unsupervised learning method like K-Means and Artificial Neural Network (ANN) for prediction of heart diseases. Krishnaiah *et al.* [12] studied many fuzzy approaches using data mining for finding heart anomalies.

Takci [13] investigated on the importance of feature selection with respect to diagnosis of heart diseases accurately. Ismail *et al.* [14] investigated on big data analytics for prediction of heart diseases. Buettner and Schunter [15] used different machine learning methods for disease prediction. Supervised machine learning techniques are used in [16], [17] and [26] for detecting heart diseases. Deep learning methods are explored in [20], [23] for better accuracy in heart disease prediction. A hybrid method using fuzzy logic and ANN is used in [18] while traditional risk factors and machine learning are used in [19] for finding cardiovascular mortality.

The notion of real time prediction and cardiac health monitoring are studied in [20] and [21] respectively. Automatic detection of heart disease with machine learning [24] and big data analytics for cardiovascular disease analysis are other important researches found. A decision support system for cardiac health monitoring is studied in [22]. From the literature, it is found that there are unsupervised and supervised learning methods widely used for predicting heart diseases. However, there is need for having a framework that combines and reaps benefits of both supervised and unsupervised learning methods. This gap is considered in this paper.

3. Proposed framework

We proposed a framework for efficient detection of heart diseases using data-driven ML approach which is useful for CDSS. Unlike traditional approaches, it exploits the power of unsupervised learning methods prior to the use of classification techniques as prediction models. Thus the proposed framework is hybrid in nature. It is named as Hybrid Machine Learning for Heart Disease Prediction (HML-HDP). K-Means and DBSCAN are widely used clustering (unsupervised learning) methods. In the proposed framework, these techniques are used for clustering the data prior to using it for prediction of heart diseases. Entropy and Gain based Feature Selection (EGFS) proposed in our prior work [27] is used to select appropriate features. After performing clustering with the aforementioned algorithms, prediction process is completed with the widely used classification algorithms known as SVM, NB and RF. Equipped with the knowledge of clustering and feature selection, these prediction models are leveraged to have better performance over the baseline classification algorithms. This is an important aspect in the proposed framework.

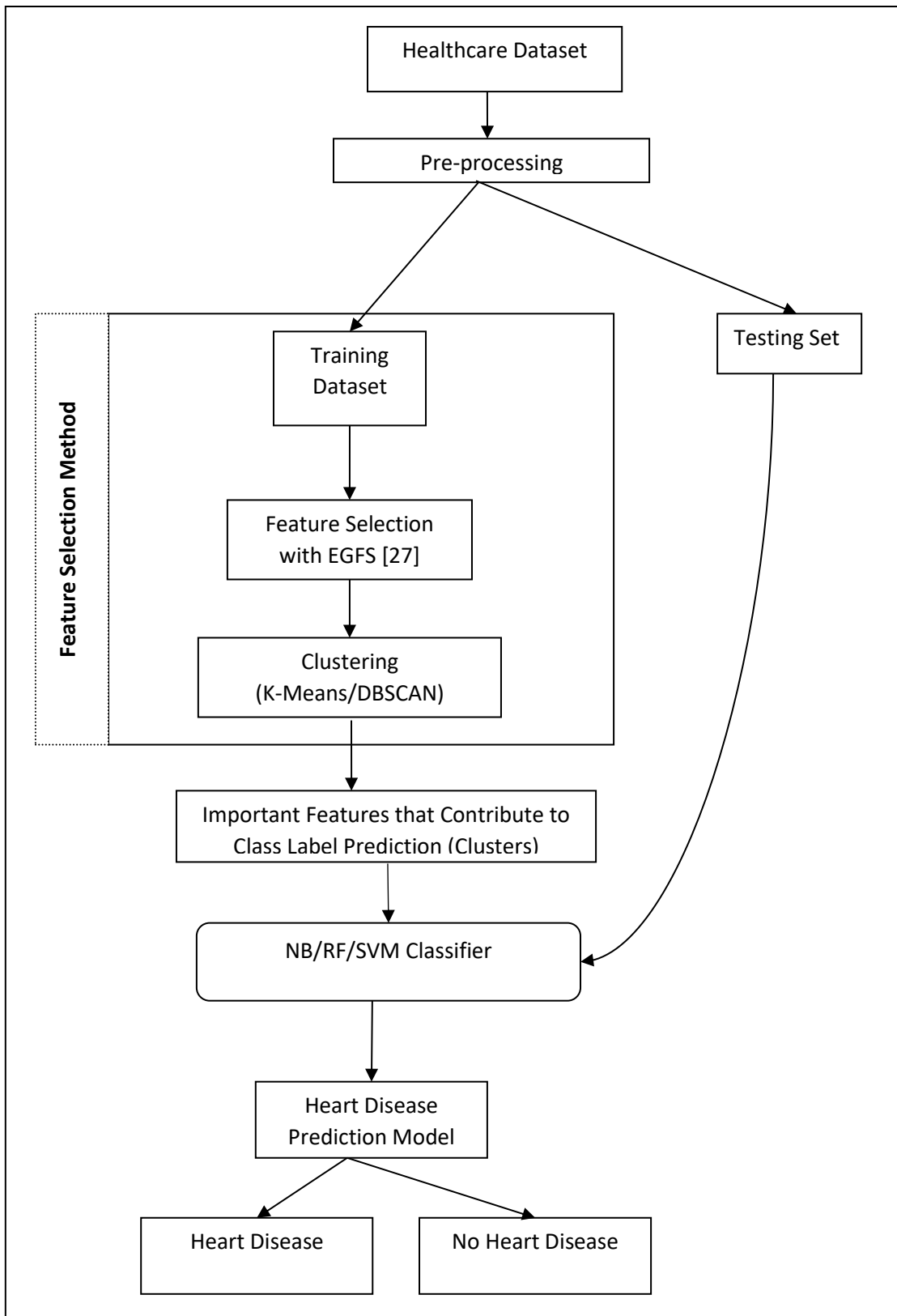


Figure 1: A hybrid framework for heart disease prediction

As presented in Figure 1, the framework healthcare dataset taken from UCI repository is used for experiments. The dataset is subjected to general pre-processing and then the data is divided into 80% training set and 20% testing set. Then the training data is subjected to feature selection using feature selection method proposed by us in [27]. Afterwards, the training data associated with selected features is used for unsupervised learning. The results of

clustering are used for supervised learning methods such as SVM, NB and RF for learning and prediction. Naïve Bayes is one of the widely used classifiers. It is a probabilistic machine learning model that's used for classification task. The crux of the classifier is that it is based on the Bayes theorem. It is used in many real world applications for classification. For instance, it is used in heart disease prediction, spam filtering, classifying cancer diseases, segmenting documents and predicting sentiments in online reviews. The Naïve Bayes classification technique is based on the Bayes theory. The features it uses are independent and hence the name naïve. It does mean that when a value of a feature is changed, it does not affect other feature directly. This algorithm is found faster as it is probabilistic. It is also scalable in nature and suitable for applications where scalability is in demand. It has its associated concepts like Bayes Rule and conditional probability. The Bayes theorem is as in Eq. 1.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

Using Bayes theorem, we can find the probability of **A** happening, given that **B** has occurred. Here, **B** is the evidence and **A** is the hypothesis. The assumption made here is that the features are independent. That is presence of one particular feature does not affect the other. Considering playing golf problem, for example, we classify whether the day is suitable for playing golf, given the features of the day. The columns represent these features and the rows represent individual entries. If we take the first row of the dataset, we can observe that is not suitable for playing golf if the outlook is rainy, temperature is hot, humidity is high and it is not windy. We make two assumptions here, one as stated above we consider that these predictors are independent. That is, if the temperature is hot, it does not necessarily mean that the humidity is high. Another assumption made here is that all the predictors have an equal effect on the outcome. That is, the day being windy does not have more importance in deciding to play golf or not. According to this example, Bayes theorem can be rewritten as in Eq. 2.

$$P(y|X) = \frac{P(X|y)P(y)}{P(X)} \quad (2)$$

The variable **y** is the class variable (play golf), which represents if it is suitable to play golf or not given the conditions. Variable **X** represent the parameters/features. It is computed as in Eq. 3.

$$X = (X_1, X_2, X_3, \dots, X_N) \quad (3)$$

present the features, i.e. they can be mapped to outlook, temperature, humidity and windy. By substituting for **X** and expanding using the chain rule we get the proposition as in Eq. 4.

$$P(y|X_1, \dots, X_n) = \frac{P(X_1|y)P(X_2|y)\dots P(X_n|y)P(y)}{P(x_1)P(x_2)\dots P(x_n)} \quad (4)$$

Now, you can obtain the values for each by looking at the dataset and substitute them into the equation. For all entries in the dataset, the denominator does not change, it remains static. Therefore, the denominator can be removed and a proportionality can be introduced as in Eq. 5.

$$P(y|X_1, \dots, X_n) \propto P(y) \prod_{i=1}^n P(X_i|y) \quad (5)$$

In this case, the class variable(**y**) has only two outcomes, yes or no. There could be cases where the classification could be multivariate. Therefore, we need to find the class **y** with maximum probability as in Eq. 6.

$$y = \operatorname{argmax}_y P(y) \prod_{i=1}^n P(X_i|y) \quad (6)$$

Using the above function, we can obtain the class, given the predictors.

Support Vector Machine (SVM) is one of the widely used machine learning techniques used for prediction purposes like heart disease prediction. It can be used in many real world applications like hand-written character recognition, cancer prediction, protein classification, image classification and text categorization to mention few. It is a discriminative classifier which predicts class labels based on training knowledge. It achieves the classification formally with a definition of hyperplane. Thus it can provide largest minimum distance known as maximum margin associated with training samples. With this SVM can minimize the margin of the data used for training.

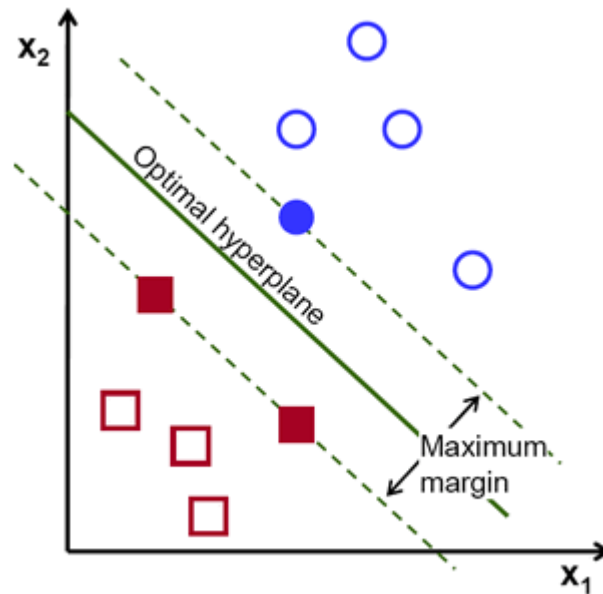
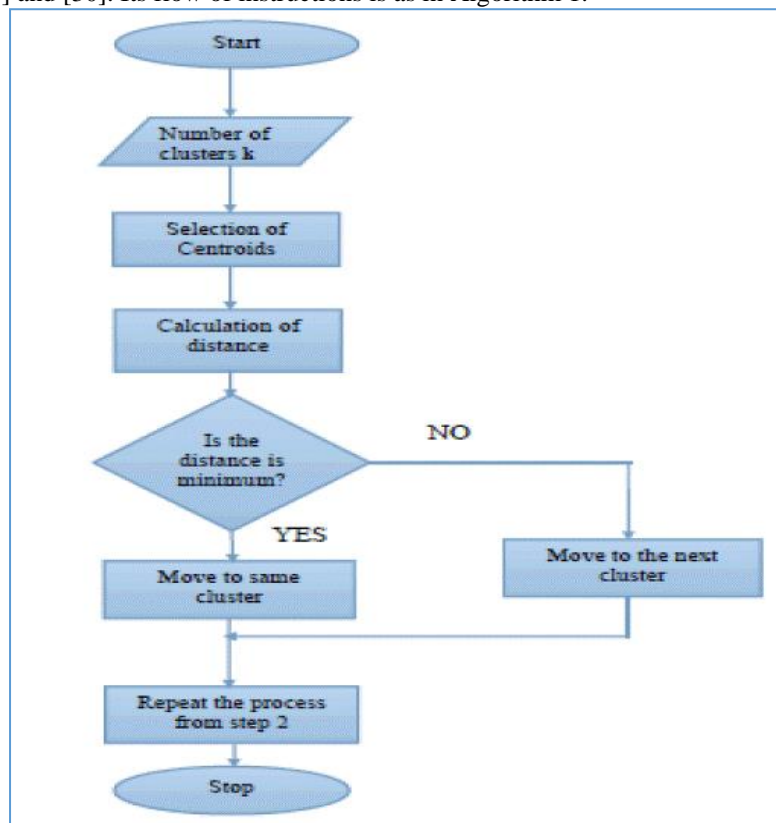


Figure 2: Illustrates how hyperplane is formed for discrimination

The optimal hyperplane with maximum margin is able to separate or discriminate two classes. In many research papers, SVM is found to have superior performance over its counterparts as found in the literature. It is used as part of the proposed methodology in this research. It makes use of the hyperplane illustrated in Figure 2. SVM is basically a binary classifier which contains two class labels. In order to have multiple class labels, it is used with appropriate kernel parameter. With respect to clustering, K-Means is the widely used clustering algorithm. It needs k value as input and divides the given data into k number of clusters. It is used in spam detection research as well discussed in [28], [29] and [30]. Its flow of instructions is as in Algorithm 1.



Algorithm 1: K-Means algorithm

The algorithm takes dataset and number of clusters as input. Then it selects the centroids and computes distance. Based on the distance, it makes clustering decisions. This process is continued with adjustment in centroids based on the distance measure. The final clustering results in given number of clusters. DBSCAN [31] is another popular clustering algorithm. It is based on the density of dataset provided. It performs clusters based on the underlying data. Its algorithm is as follows.

1. Choose a data point randomly
2. Use parameters like MinPts and Eps in order to reclaim possible data with density reachable from the selected point.
3. If the chosen point is core point, it satisfies Eps and MinPts
4. If the chosen point is a border point, then data is not density-reachable from the selected point.
5. This process is repeated until the convergence.

Algorithm 2: Steps in DBSCAN algorithm

As presented in Algorithm 2, DBSCAN algorithm performs clustering based on two parameter values such as Eps and MinPts.

4. Evaluation methodology

Evaluation of the proposed spam detection models (baseline and hybrid) is made using measures derived from confusion matrix. Table 1 shows the confusion matrix.

	Ground Truth (Yes)	Ground Truth (No)
Result of proposed detection model (Yes)	True Positive (TP)	False Positive (FP)
Result of proposed detection model (No)	False Negative (FN)	True Negative (TN)

Table 1: Confusion matrix

As per the confusion matrix, ground truth is the actual correct prediction value either positive (yes) or negative (no). Based on the comparison between prediction results and actual ground truth results, there are four important cases such as TP, FP, FN and TN. Out of them TP and TN are correct predictions while FN and FP are wrong predictions.

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

$$F - Measure = 2 * \frac{(Precision * Recall)}{(Precision + Recall)} \tag{9}$$

As shown in Eq. 1, precision reflects exactness of the results or quality of results while the recall (Eq. 8) shows completeness. The harmonic mean of precision and recall is known as F-measure or F-score which is provided in Eq. 9.

5. Experimental setup

A prototype application is built using Python data science platform for empirical study. A PC with 4 GB of RAM, 2.40 GHz processor from Intel and Windows 10 operating system is used for experiments. Spyder is the IDE used for development of the application. It is part of Anaconda distribution of Python.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	id	Age	Sex	ChestPain	RestBP	Chol	Fbs	RestECG	MaxHR	ExAng	Oldpeak	Slope	Ca	Thal	AHD
2	1	67	1	4	160	286	0	2	108	1	1.5	2	3	3	2
3	2	67	1	4	120	229	0	2	129	1	2.6	2	2	7	1
4	3	37	1	3	130	250	0	0	187	0	3.5	3	0	3	0
5	4	41	0	2	130	204	0	2	172	0	1.4	1	0	3	0
6	5	56	1	2	120	236	0	0	178	0	0.8	1	0	3	0
7	6	62	0	4	140	268	0	2	160	0	3.6	3	2	3	3
8	7	57	0	4	120	354	0	0	163	1	0.6	1	0	3	0
9	8	63	1	4	130	254	0	2	147	0	1.4	2	1	7	2
10	9	53	1	4	140	203	1	2	155	1	3.1	3	0	7	1
11	10	57	1	4	140	192	0	0	148	0	0.4	2	0	6	0
12	11	56	0	2	140	294	0	2	153	0	1.3	2	0	3	0
13	12	56	1	3	130	256	1	2	142	1	0.6	2	1	6	2
14	13	44	1	2	120	263	0	0	173	0	0	1	0	7	0

Figure 3: An excerpt from training set

As presented in Figure 3, It is observed that there are 15 columns in the dataset given for training. The dataset is made up of patients' vital signs for heart diseases prediction. The last attribute is the class label. It does mean that a training set needs to have class label for every instance.

6. Experimental results

Experiments are made with the aforementioned dataset. Two sets of experiments are made. In the first set of experiments, only supervised machine learning method along with feature selection is used for heart disease prediction. In the second set of experiments, in order to ascertain the impact of unsupervised learning on the prediction models, K-Means and DBSCAN clustering algorithms are used along with three supervised learning methods such as NB, SVM and RF. The performance of the models is observed in terms of confusion matrix. The results are presented in terms of precision, recall and F-Measure.

Table 2: Shows heart disease prediction performance with Naïve Bayes and its hybrid counterparts

Detection Models	Performance (%)		
	Precision	Recall	F-Measure
Naïve Bayes	0.91	0.89	0.899888889
Naïve Bayes with K-Means	0.95	0.9	0.924324324
Naïve Bayes with DBSCAN	0.97	0.92	0.944338624

As presented in Table 2, the performance of the prediction models is provided in terms of precision, recall and F-Measure. The results of individual classifier like NB are compared with that of hybrids made up of NB with K-Means and NB with DBSCAN.

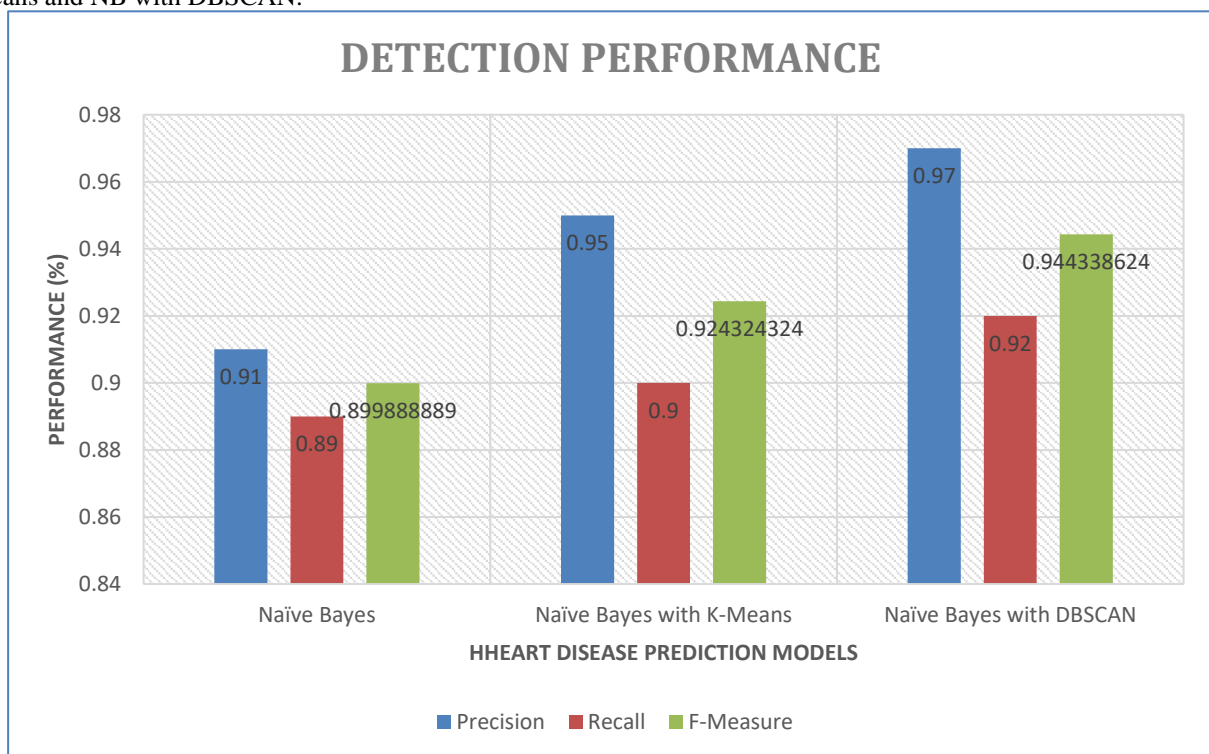


Figure 4: Performance of NB and its hybrid counterparts

As presented in Figure 4, the heart disease prediction performance of different models is compared. Prediction models are provided in horizontal axis and the vertical axis shows the performance of the models in terms of precision, recall and F-Measure. The prediction model made up of NB is compared with its hybrid counterparts like NB with K-Means and NB with DBSCAN. NB with clustering methods showed better performance over individual NB. However, the NB with DBSCAN outperformed NB with K-Means. From the results, therefore, it is understood that DBSCAN has revealed better performance over K-Means as far as unsupervised learning is

considered. NB with DBSCAN showed Highest performance with precision 0.97, recall 0.92 and F-Measure with 0.94433.

Table 3: Shows heart disease prediction performance with SVM and its hybrid counterparts

Detection Models	Performance (%)		
	Precision	Recall	F-Measure
Support Vector Machine	0.88	0.88	0.88
SVM with K-Means	0.94	0.91	0.924756757
SVM with DBSCAN	0.96	0.93	0.944761905

As presented in Table 3, the performance of the prediction models is provided in terms of precision, recall and F-Measure. The results of individual classifier like SVM are compared with that of hybrids made up of SVM with K-Means and SVM with DBSCAN.

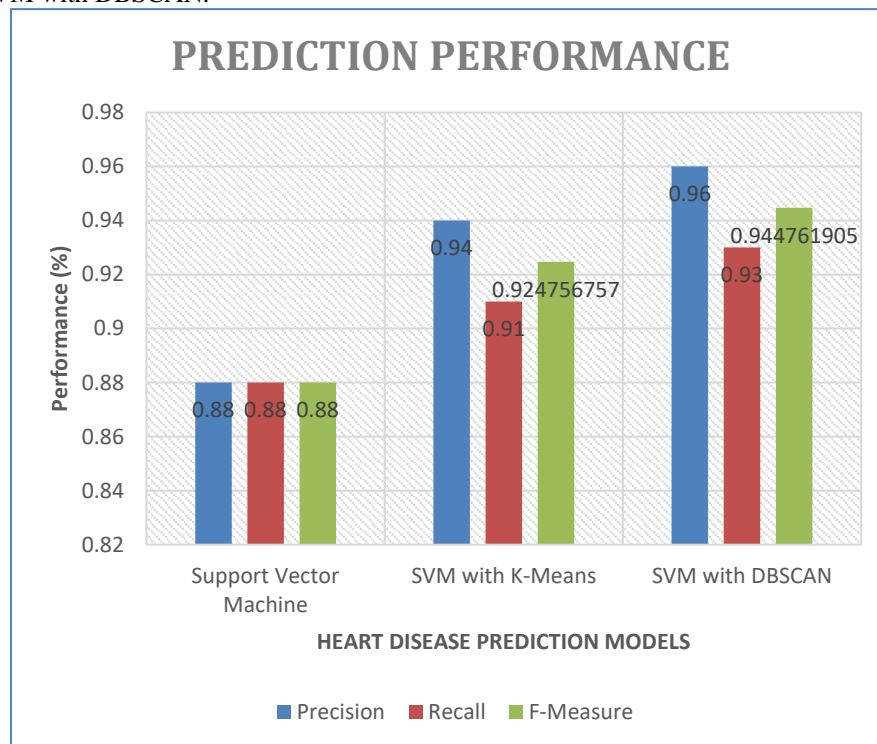


Figure 5: Performance of NB and its hybrid counterparts

As presented in Figure 5, the heart disease prediction performance of different models is compared. Prediction models are provided in horizontal axis and the vertical axis shows the performance of the models in terms of precision, recall and F-Measure. The prediction model made up of SVM is compared with its hybrid counterparts like SVM with K-Means and SVM with DBSCAN. SVM with clustering methods showed better performance over individual SVM. However, the SVM with DBSCAN outperformed SVM with K-Means. From the results, therefore, it is understood that DBSCAN has revealed better performance over K-Means as far as unsupervised learning is considered. SVM with DBSCAN showed high performance with precision 0.96, recall 0.93 and F-Measure with 0.9447.

Table 4: Shows heart disease prediction performance with RF and its hybrid counterparts

Detection Models	Performance (%)		
	Precision	Recall	F-Measure
Random Forest	0.85	0.85	0.85
Random Forest with K-Means	0.91	0.88	0.894748603
Random Forest with DBSCAN	0.93	0.91	0.919891304

As presented in Table 4, the performance of the prediction models is provided in terms of precision, recall and F-Measure. The results of individual classifier like RF are compared with that of hybrids made up of RF with K-Means and RF with DBSCAN.

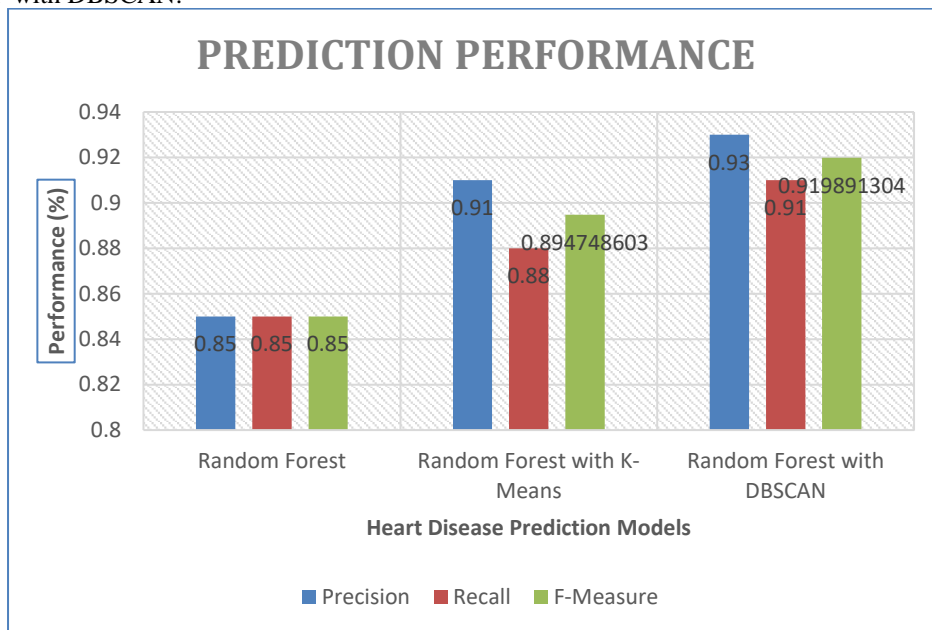


Figure 6: Performance of NB and its hybrid counterparts

As presented in Figure 6, the heart disease prediction performance of different models is compared. Prediction models are provided in horizontal axis and the vertical axis shows the performance of the models in terms of precision, recall and F-Measure. The prediction model made up of RF is compared with its hybrid counterparts like RF with K-Means and RF with DBSCAN. RF with clustering methods showed better performance over individual RF. However, the RF with DBSCAN outperformed RF with K-Means. From the results, therefore, it is understood that DBSCAN has revealed better performance over K-Means as far as unsupervised learning is considered. RF with DBSCAN showed high performance with precision 0.96, recall 0.93 and F-Measure with 0.9447.

7. Conclusion and future work

We proposed a framework known as Hybrid Machine Learning for Heart Disease Prediction (HML-HDP) for heart disease prediction. It exploits the power of unsupervised learning prior to supervised learning. Unlike traditional approaches, it tries to have clusters of data to optimize performance of prediction models. Two sets of experiments are made. In the first set of experiments, only supervised machine learning method along with feature selection is used for heart disease prediction. In the second set of experiments, in order to ascertain the impact of unsupervised learning on the prediction models, K-Means and DBSCAN clustering algorithms are used along with three supervised learning methods such as NB, SVM and RF. The results of the hybrid models are compared with that of individual prediction models. SVM with DBSCAN showed Highest performance with precision 0.96, recall 0.93 and F-Measure with 0.9447. SVM as an individual prediction model exhibited 0.88 as precision, 0.88 recall and 0.88 F-Measure. Therefore, from the results, it is ascertained that clustering has significant impact on the performance of heart disease prediction models. Another observation is that SVM hybrid with DBSCAN is better than its K-Means counterpart. In future, it is useful to investigate further on the clustering algorithms for further improvement of the proposed framework. Towards this end, an ensemble of clustering approaches is to be investigated.

References

- A. Shadi A. Aljawarneh, Reem Jaradat, Abdel Salam Maatuk and Abdullah Alhaj . (2016). Gene Profile Classification: A Proposed Solution for Predicting Possible Diseases and Initial Results. IEEE , p1-8.
- B. Jagdeep Singh , Amit Kamra , Harbhag Singh. (2016). Prediction of Heart Diseases Using Associative Classification. IEEE, p1-8.

- C. Milad Zafar Nezhad , Dongxiao Zhu, , Najibesadat Sadati , Kai Yang , Phillip Levy. (2017). SUBIC: A Supervised Bi-Clustering Approach for Precision Medicine. *IEEE*. 1, p1-8.
- D. Panagiotis Antiochos, , Pedro Marques-Vidal, , Aaron McDaid, , Gerard Waeber, , Peter Vollenweider. (2015). Association between parental history and genetic risk scores for coronary heart disease prediction: The population-based CoLaus study. *Elsevier*, p1-8.
- E. Indu Yekkala, Sunanda Dixit, M.A.jabbar . (2017). Prediction of heart disease using Ensemble Learning and Particle Swarm Optimization. *IEEE*, p1-9.
- F. Rajat Mittal, Jung Hee Seo , Vijay Vedula , Young J. Choi , Hang Liu , H. Howie Huang , Saurabh Jain , Laurent Younes , Theodore Abraham , and Richard T. George. (2016). Computational Modeling of Cardiac Hemodynamics: Current Status and Future Outlook . *Elsevier*. 305 , p1-27.
- G. Priyan Malarvizhi Kumar , Usha Devi Gandhi. (2017). A novel three-tier Internet of Things architecture with machine learning algorithm for early detection of heart diseases. *Elsevier*, p1-14.
- H. Animesh Kumar Paul , Pintu Chandra Shill, Md. Rafiqul Islam Rabin, M. A. H. Akhand .(2016). Genetic Algorithm Based Fuzzy Decision Support System for the Diagnosis of Heart Disease . *IEEE*, p1-7.
- I. František Babič, Jaroslav Olejár, Zuzana Vantová, Ján Paralič. (2017). Predictive and Descriptive Analysis for Heart Disease Diagnosis. *IEEE*, p1-9.
- J. Prabal Verma and Sandeep K.Sood. (2017). Cloud-centric IoT based disease diagnosis healthcare framework. *Elsevier*, p1-12.
- K. Amita Malav, Kalyani Kadam, Pooja Kamat. (2017). PREDICTION OF HEART DISEASE USING K-MEANS and ARTIFICIAL NEURAL NETWORK as HYBRID APPROACH to IMPROVE ACCURACY . *researchgate*. 9 (4), p1-6.
- L. V. Krishnaiah, G. Narsimha, PhD, N. Subhash Chandra, PhD. (2016). Heart Disease Prediction System using Data Mining Techniques and Intelligent Fuzzy Approach: A Review. *Citeseer*. 136 (2), p1-9.
- M. Hidayet TAKCI. (2017). Improvement of heart attack prediction by the feature selection methods. *TUB İTAK*. . (.), p1-10.
- N. AHMED ISMAIL, SAMIR ABDLERAZEK, I. M. EL-HENAWY . (2020). BIG DATA ANALYTICS IN HEART DISEASES . *JATIT*. 98 (11), p1-12.
- O. Ricardo Buettner and Marc Schunter. (2019). Efficient machine learning based detection of heart disease. *IEEE*, p1-6
- P. Manikandan, A., Suganya, K., Saranya, N., Sudha, V., & Sweetha, S. (2017). Assessment of Intracardiac Masses Classification. *Journal of Chemical and Pharmaceutical Sciences*, 5, 101–103.
- Q. Purushottama, Prof. (Dr.) Kanak Saxena, Richa Sharma . (2016). Efficient Heart Disease Prediction System. *Elsevier*, p1-8.
- R. Oluwarotimi Williams Samuel, Grace Mojisola Asogbona, Arun Kumar Sangaihc, Peng Fanga, Guanglin Li. (2016). An integrated decision support system based on ANN and Fuzzy_AHP for heart failure risk prediction. *Elsevier*, p1-11.
- S. Manikandan, A., & Jamuna, V. (2017). Single Image Super Resolution via FRI Reconstruction Method. *Journal of Advanced Research in Dynamical and Control Systems*, 9(2), 23–28
- T. Michael Tschannen, Thomas Kramer, Gian Marti, Matthias Heinzmann, Thomas Wiatowski. (2016). Heart Sound Classification Using Deep Structured Features. *IEEE*, p1-4.
- U. Abdur Rahim Mohammad Forkana, Ibrahim Khalil, Mohammed Atiquzzaman. (2017). ViSiBiD: A learning model for early discovery and real-time prediction of severe clinical events using vital signs as big data. *researchgate*, p1-15.
- V. Shuroq Hijazi and Alex Page, Burak Kantarci, Burak Kantarci. (2016). Machine Learning in Cardiac Health Monitoring and Decision Support. *tolgasoyata*, p1-11.
- W. Geert Litjens, Francesco Ciompi, Jelmer M. Wolterink, Bob D. de Vos, Tim Leiner, Jonas Teuwen, Ivana Išgum. (2019). State-of-the-Art Deep Learning in Cardiovascular Image Analysis. *Elsevier*. 12 (8), p1-17.
- X. Manikandan, A., & Sakthivel, J. (2017a). Recognizable Proof of Biometric System With Even Distorted And Rectification States. *Journal of Advanced Research in Dynamical and Control Systems*, 9(2), 1393–1398.
- Y. PRASAN KUMAR SAHOO, SUVENDU KUMAR MOHAPATRA AND SHIH-LIN WU. (2016). Analyzing Healthcare Big Data With Prediction for Future Health Condition. *IEEE*. 4 , p1-14.
- Z. Alexander R. van Rosendael , Gabriel Maliakal , Kranthi K. Kolli , Ashley Beecy , Subhi J. Al’Aref , Aeshita Dwivedi , Gurpreet Singh , Mohit Panday , Amit Kumar , Xiaoyue Ma , Stephan Achenbach , Mou. (2018). Maximization of the usage of coronary CTA derived plaque information using a machine learning based algorithm to improve risk stratification; insights from the CONFIRM registry. *Elsevier*, p1-7.

- AA. Pulugu Dileep, Kunjam Nageswara Rao, Prajna Bodapati, “An Efficient Feature Selection Based Heart Disease Prediction Model”, *International Journal of Advanced Science and Technology*, ISSN: 2005-4238, Vol. 28, No. 9, pp. 309-323.
- BB. Wang, S., Zhang, X., Cheng, Y., Jiang, F., Yu, W., & Peng, J. (2018). A Fast Content-Based Spam Filtering Algorithm with Fuzzy-SVM and K-means. *2018 IEEE International Conference on Big Data and Smart Computing (BigComp)*. P1-7.
- CC. Bellin Ribeiro, P., Alexandre da Silva, L., & Pontara da Costa, K. A. (2015). Spam intrusion detection in computer networks using intelligent techniques. *2015 IFIP/IEEE International Symposium on Integrated Network Management (IM)*. P1-4.
- DD. Banitaan, S., Nassif, A. B., & Azzeh, M. (2015). Class Decomposition Using K-Means and Hierarchical Clustering. *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*. P1-5.
- EE. Dixon, B. (2017). Investigating clustering algorithm DBSCAN to self select locations for power based malicious code detection on smartphones. *2017 Third International Conference on Mobile and Secure Services (MobiSecServ)*. P1-7.