ELIMINATION OF IRRELEVANT FEATURES AND HEART DISEASE RECOGNITION BY EMPLOYING MACHINE LEARNING ALGORITHMS USING CLINICAL DATA

HUAIYU WEN[1], SUFANG LI[1], AMIN UL HAQ[2], JIAN PING LI[2], RAJESH KUMAR [2], ABDUS SABOOR[2], TAO JIANG[2]

[1]College of Computer Science, Chengdu University, Chengdu, China, 616000

[2]School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

E-MAIL: wenhuaiyu@163.com, *khan.amin50@yahoo.com, jpli2222@uestc.edu.cn, rajakumarlohano@gmail.com, sabbor_khan97@yahoo.com, tao1024@yahoo.com

Abstract:

A heart disease diagnosis method has been proposed for effective heart disease diagnosis. In the proposed method Machine Learning (ML) classifiers have been used for detection of heart disease. Chi square feature selection algorithm has been used for related feature selection to improve the prediction performance of machine learning models. Cross validation, method Hold out has been employed for model hyper parameters tuning and best model selection. Furthermore, performance evaluation metrics, such as classification accuracy, specificity, sensitivity, Matthews' correlation coefficient and execution time have been used for model performance evaluation. The Cleveland heart disease data set has been used for testing of the proposed method. The experimental results demonstrated that proposed method has achieved high performance as compared to state of the art methods. Furthermore, the proposed method performance has been compared with deep learning model. Thus, the proposed method will support the medical professional to diagnosis heart disease efficiently and could easily incorporated in healthcare for diagnosis of heart disease.

Keywords:

Machine learning algorithms; Heart disease; diagnosis; accuracy; Feature selection

1. Introduction

Heart disease is a great medical issue and numerous people suffered from it around the world [1]. The HD rate in the United States (US) are more [2]. The signs of HD patient contains, rapidity of inhalation, physical body weakness, swollen of feet [3]. The examination procedures that are used to identify HD were very difficult and its resulting difficulty [4]. The diagnosis and treatment of HD is difficult, particularly in the poor countries [5]. The proper diagnosis of the HD can reduced the risk of HD [6]. According to the report of European Society of Cardiology that 26 million children universal were identified with HD [7]. To the identifying HD through invasive based methods are founded on the examination of patient's health history and analysis of related signs by health specialists. These methods are not suitable for HD diagnosis [8].

To solve the problems in invasive based HD method, a non-invasive method base on machine learning and rough set [9, 10] have been designed by experts for HD detection [11]. Cleveland HD dataset is online available on the UCI which was used by many researchers [12, 13].

In [13] recommended a HD detection approach employing logistic regression and SVM and attained high results. In [14] authors developed a method with global evolutionary methods and attained high prediction accuracy. Furthermore, FS method was incorporated for feature selection and the proposed model gained high performance on selected features. Gudadhe et al. [15] used MLP and SVM for HD identification and 80.41% accuracy obtained. Humar ,et al. [16] developed a HD detection method employing Neural Network that integrates a Fuzzy Neural Network and 87.4% accuracy obtained. In [17] Proposed a heart disease method by employing ML algorithms and achieved good performance.

In this study a new HD diagnosis method has been proposed. Feature selection algorithm Chi squared was used for related feature selection to improve the performance of the classifier from Cleveland data. Five machine learning models were used to predict HD. Model evaluation metrics have employed for model evaluation. The performance proposed method was compared to state of the art method.

The major contribution of the proposed study as follow:

1. Chi squared FS algorithm was used for feature selection to increase the prediction accuracy of the classifiers
2. The proposed method achieved high accuracy on selection features set.
3. The proposed method effectively diagnosis of HD and can be easily incorporated in health care.

The paper organization in such a way that I section proposed method methodology has been explained. The experiments and results discussion are being reported in 3 section. In section 4 conclusions and future work are discussed.

2. Materials and method

The data set and background of the proposed has been presented in below sections in details.

2.1. Data Set

The "Cleveland heart disease dataset 2016" has been used by different scholars [13]. This data set On UCI repository is available.

2.2. Feature Selection Based on Chi-squared Algorithm

Chi Squared is FS algorithm and in various research, study used it for FS [27] is algorithm calculates X^2 between each feature and target variable and select the necessary number of features with the best X^2 scores. The foundation is that if a feature is autonomous of the target variable it is uninformative for classifying clarifications. The chi-square is stated mathematically in eq (1).

$$X^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \quad (1)$$

Where O_i is the number of interpretations in class and E_i is the number of probable explanations in class i if there is no relation between the feature and target variable.

2.3. Classification models

The following ML algorithms are incorporated in classification process of the proposed method.

2.3.1. Logistic regression

One of the famous ML model is Logistic Regression (LR) [18] and it is incorporated mostly in classification related problems [21, 22, 25, 26]. It is used for binary machine learning problem of classification to predict the value of predictive variable y when y$\in$ {0, 1}.

The LR logistic function can be expressed in eq (2).

$$\mathrm{h}\theta(\mathrm{x}) = \mathbf{g}(\theta^{\mathrm{T}}\mathrm{x}) \quad (2)$$

Where, $\mathbf{g}\ (\mathbf{z}) = \frac{\mathbf{1}}{\mathbf{1}+x^{-z}} \mathbf{h\theta(x)}\ = \frac{\mathbf{1}}{\mathbf{1}+x^{-z}}$

The Objective function of LR expressed mathematically in eq (3).

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} Cost(h\,\theta(\,x^{(i)}), y^{(i)}) \quad (3)$$

2.3.2. Decision Tree

Decision Tree is machine learning model [19] and used for problems of classification [20, 21]. Tree models split the dataset numerous epochs according to definite features limit values. Over piercing, various subclasses of the data-set are design, with each sample fitting to one sub-set. The last sub-sets are terminal (leaf) nodes while the in-between sub-sets are interior nodes. To guess the outcome in each terminal node, the systematic outcome of the training-data in this node is incorporated.

2.3.3. Support Vector Machine (SVM)

For binary classification related problem SVM is usually employed [23, 24]. In the situation of binary combination, the instances are divided by a hyper-plane $w^T x + b = 0$, and w and d-dimensional co-efficient vectors, which is group for the external hyper-plane. The b, are offset from the origin, x is values of the data-set. SVM gets w and b outcomes. The w can resolve in the linear circumstance by calculation Lagrangian multipliers. The w result written as w=, $\sum_{i=1}^{n} \alpha_i y_i x_i$, the n is the number of supported vectors, yi is the output target labels cross ponding to x. The w and b are calculated as written in eq (3). The linear- discriminating function written in eq (4).

$$g(x) = \mathrm{sgn}\left(\sum_{i=1}^{n} \alpha_i y_i\, x_i^T x + b\right) \quad (4)$$

The nonlinear setup expressed in eq (5) for Kernel Trick (KT) and Decision Function.

$$\mathrm{g(x)} = \mathrm{sgn}\left(\sum_{i=1}^{n} \alpha_i y_i K(x_i, x) + \mathrm{b}\right) \quad (5)$$

The Positive Semidefinite Functions (PSF) that follow Mercer's condition as Kernel Functions (KF) [8].

2.3.4. Naïve Bayes (NB)

The NB is ML algorithm and employing mostly for problems of classification [28, 29].

and it is built on a constrained probability statement to state the class of a new vector of features. The NB procedures the training data to find out the restricted possibility charge of vectors for a specified class.

2.4. Hold out cross validation

The hold out cross validation method has been for training and validation of the model. According to hold out CV 70% of the data used for training and 30% for validation of the model.

2.5. Performance evaluation metrics

Performance evaluation metrics were used for model performance evaluation. These metrics are computed with the help of the confusion matrix. Table 1 shows the binary classification matrix.

Table 1 Confusion matrix [30, 31, 32]

	Predicted Absence of HD subject (0)	Predicted presence of HD subject (1)
Actual Healthy Subject 0	TN	FP
Actual HD Subject 1	FN	TP

From Table 1 we computed the following performance evaluation metrics and mathematically shown in eqs (6), (7), (8), (9), (11), and (11) respectively.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \text{ x100\%} \quad (6)$$

$$\text{Sensitivity}= \frac{TP}{TP+FN}\text{x100\%} \quad (7)$$

$$\text{Specificity} = \frac{TN}{TN+FP} \text{ x100\%} \quad (8)$$

$$\text{Precision} = \frac{Tp}{Tp+FP} \text{ x100\%} \quad (9)$$

$$\text{MCC}=\frac{TP\times TN-FP\times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \text{ x100\%} \quad (10)$$

$$F1-score = 2\,.\frac{\text{precision.recall}}{\text{precision+recall}} \quad (11)$$

2.6. Proposed classification method for heart disease

In the proposed method of heart disease identification, the Cleveland heart disease data set was preprocess in order to normalized data set, Chi squared FS method was used for feature selection and then these features were used for training and testing of the models. Hold out cross validation method was used for training and testing of the models. Furthermore, performance evaluation metrics were used for model evaluation of performance. The pseudocode of the proposed method has been given in algorithm1. The proposed method Flow chart has been given in Figure 1.

Alg.1 pseudocode of proposed HD diagnosis method

Begin
Step1: The HD dataset using pre-processing methods;
Step2: Features selection using Chi squared;
Step3: Hold out method was used for model training and testing
Step4: Metrics of performance evaluation were computed for model evaluation;
Step5: Finish;
End

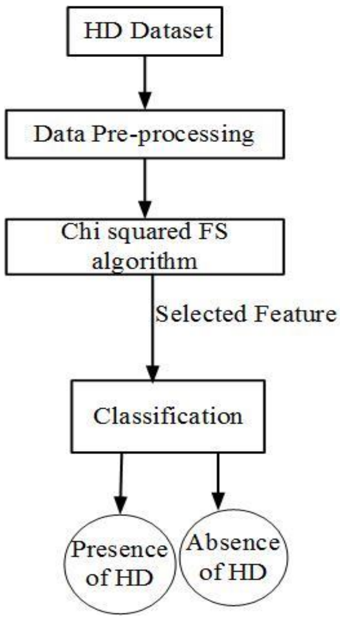

Fig.1 Flow chart of proposed HD method Experimental results and discussion

3. Experimental results and discussion

Different experiment has been performed in order to test the proposed method. Cleveland heart disease data set has been used in our experiments for testing of the proposed method. Chi squared algorithm has been used for feature selection and on these features set the classifiers have been trained and tested. Hold out cross validation method was used for models validation. Different performance evaluation metrics have computed for models perfo4mance evaluation. For simulation python programming language has been used on Intel(R) CoreTM i5-2400CPU @3.10 GHz PC window 10 with different machine learning libraries. All experimental results have been reported in tables and graphically demonstrated for better understanding.

3.1. Pre-processing of data set

The data set has been preprocess using preprocessing techniques in order to normalized. After preprocessing the data set has 297 instances with 13 features. The process data set has been used for all experiments.

Figure 3. Histogram of data set

3.2. Feature selection by Chi squared FS algorithm

After the data preprocessing, feature selection process has been performed by using Chi squared feature selection algorithm. The feature selected by Chi squared algorithm has been reported in Table 2. According to Table 2 these are important features in the prediction of HD. These selected features have been used for training and testing of the ML models. These have been labeled and the weight of these features have been shown in Table 2. The selected features by Chi square algorithm have been graphically shown in Figure 2 for better understanding.

Table 2 Features selected by Chi squared FS algorithm

S.no	Feature name	Feature code	weight
1	Sex	SEX	0.523
2	Chest pain	CPT	0.217
3	Resting blood pressure	RBP	0.165
4	Serum cholesterol	SCH	0.575
5	Resting electrocardiographic	RES	0.696
6	Maximum heart rate	MHR	0.123
7	Exercise induced angina	EIA	0.298
8	Old peak	OPK	0.561
9	ST	PES	0.574
10	vessels	VCA	0.140
11	Thallium scan	THA	0.486

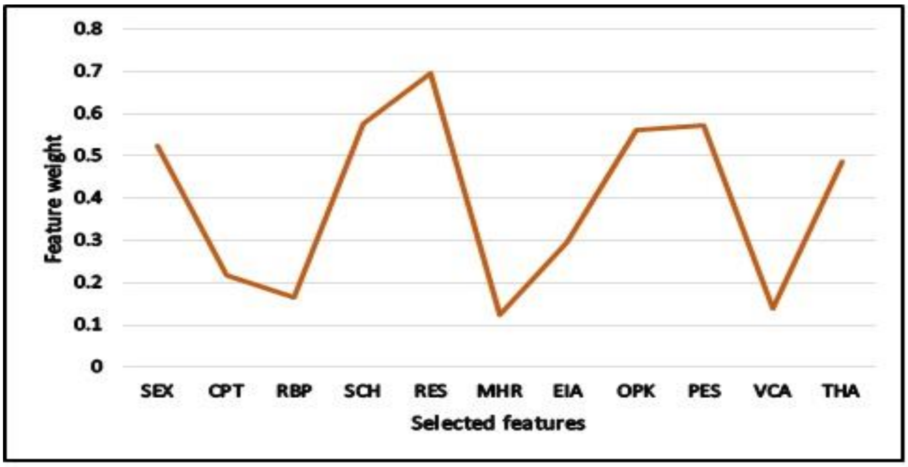

Fig.2 Feature selected by Chi squared FS algorithm

3.3. Classification results of machine learning classifiers on Full features set

The classification performance of machine learning models such as logistic regression, decision tree, support vector machine and Naïve Bayes have checked on full features set along with essential parameters of these model. All models have been trained with 70% of the data and validation with 30%. The experimental results of all models has been reported in Table 3. According to Table 3 the classifier LR achieved 85% accuracy, 87% specificity, 90% sensitivity, 86% MCC, 86% F1-score and computation of model was 0.003 seconds. DT classifier obtained these results on full features set 54% accuracy, 99% specificity, 76% sensitivity, 88% MCC, 87% F1-score and computation of model was 1.241 seconds. The classifier SVM with linear model under the hyper parameters C=1 and γ=0.001 classification results achieved on full feature set were set 86% accuracy, 89% specificity, 86% sensitivity, 87% MCC, 88% F1-score and computation of model was 1.221 seconds. Similarly, NB classification on full feature were 76% accuracy, 87% specificity, 88% sensitivity, 85% MCC, 77% F1-score and computation of model was 6.123 seconds.

The performance of SVM inn term of accuracy was high as compared to other models with full features set. The Figure 3 show the classification performance of these model with full features set. The computation of these models with full feature s set has been shown in Figure 4 for better understanding.

The computation of LR is low as compared to other models.

Table 3 Performance of models with full features set

Model	Parameter	Performances metrics					
		Acc(%)	Sp(%)	Sn(%)	MCC(%)	F1-Score(%)	Time(%)
LR	C=1	85	87	90	86	86	0.003
DT	-	54	99	76	88	87	1.241
SVM	C=1, γ=0.001	86	89	86	87	88	1.221
NB	-	76	87	88	85	77	6.123

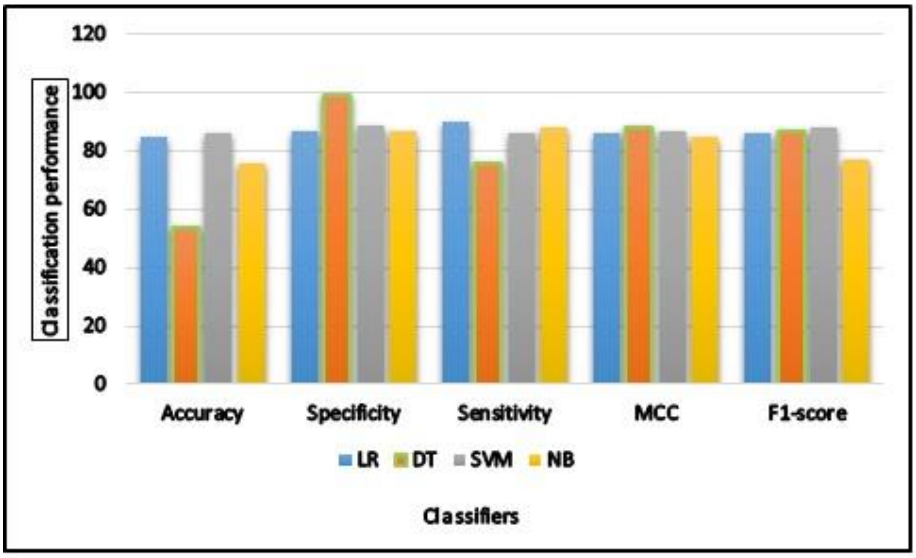

Fig.3 Classifier performance on full features set

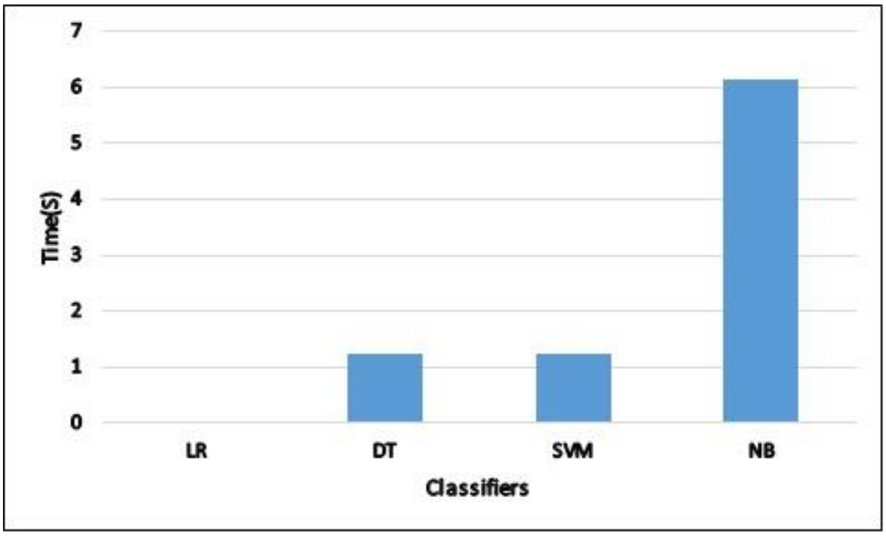

Fig.4 Processing time od models with full features set

3.4. Classification results of machine learning classifiers on Selected features set

The classification performance of machine learning models such as logistic regression, decision tree, support vector machine and Naïve Bayes have checked on selected features set along with essential parameters of these model. All models have been trained with 70% of the data and validation with 30%. The experimental results of all models has been reported in Table 4. According to Table 4 the classifier LR achieved 90% accuracy, 89% specificity, 99% sensitivity, 89% MCC, 90% F1-score and computation of model was 0.002 seconds. DT classifier obtained these results on selected features set 67% accuracy, 99% specificity, 88% sensitivity, 89% MCC, 67% F1-score and computation of model was 0.031seconds. The classifier SVM with linear model under the hyper parameters C=1 and γ=0.001 classification results achieved on selected feature set were set 93% accuracy, 99% specificity, 90% sensitivity, 90% MCC, 92% F1-score and computation of model was 0.021 seconds. Similarly, NB classification on selected feature were 78% accuracy, 80% specificity, 98% sensitivity, 89% MCC, 79% F1-score and computation of model was 2.120 seconds.

The performance of SVM inn term of accuracy was high as compared to other models on selected features set. The Figure 5 show the classification performance of these model with on selected features set. The computation of these models with selected feature s set has been shown in Figure 6 for better understanding. The computation of LR is low as compared to other models.

The best classifier SVM performance on full and selected features set has been shown in figure 7 in term of accuracy and in figure 8 in term of processing time for better understanding. According Figure 7 and 8 the classifier SVM obtained high results on selected features set as selected by Chi square FS algorithm.

Thus, features selection have important rolls in classification problem. On selected features the model performance greatly improved and computation time decreased greatly.

Table 4 Models performance on selected features set

Model	parameter	Performances metrics					
		Acc(%)	Sp(%)	Sn(%)	MCC(%)	F1-Score(%)	Time(%)
LR	C=1	90	89	99	89	90	0.002
DT	-	67	99	88	89	67	0.031
SVM	C=1, γ=0.001	93	99	90	90	92	0.021
NB	-	78	80	98	89	79	2.120

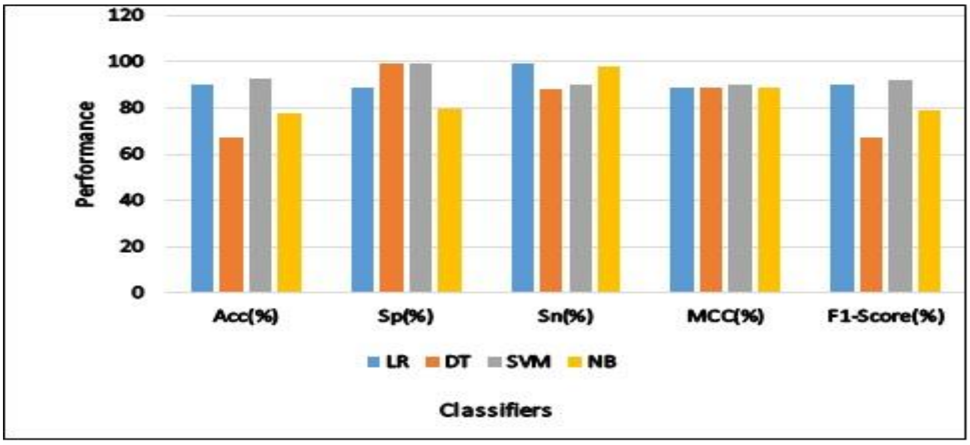

Fig.5 Models performance on selected features set

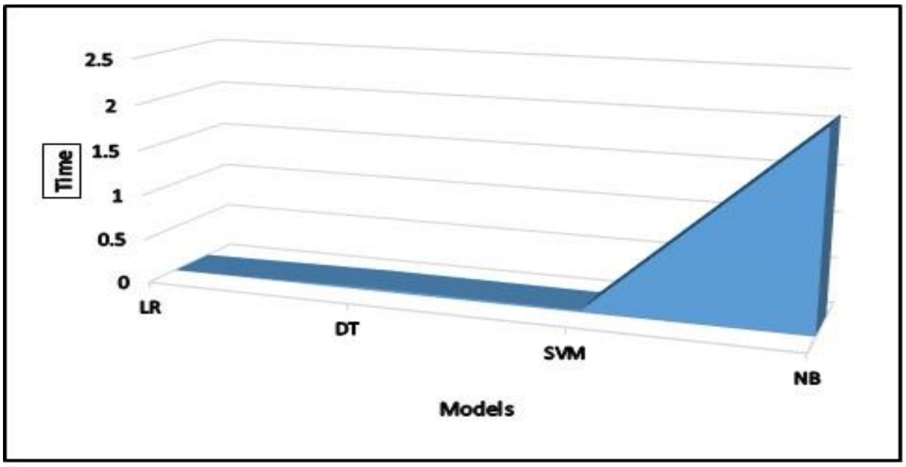

Fig.6 Processing time of models on selected features set

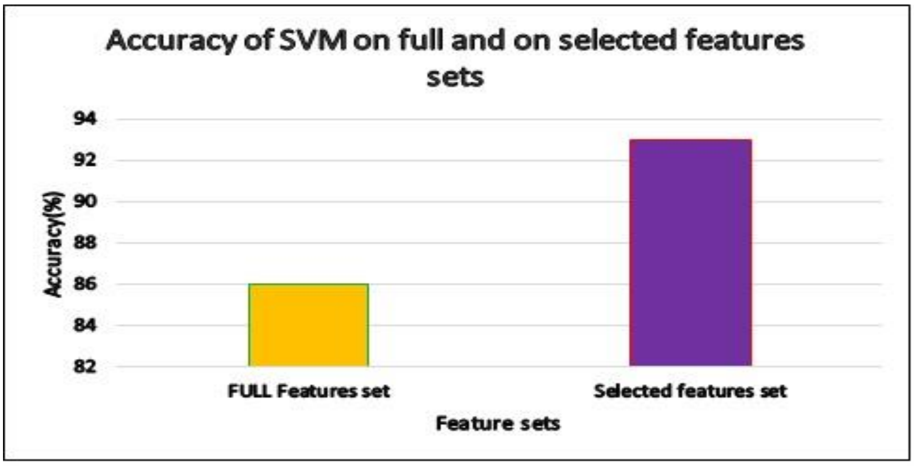

Fig.7 Accuracy of SVM on Full and on selected features sets

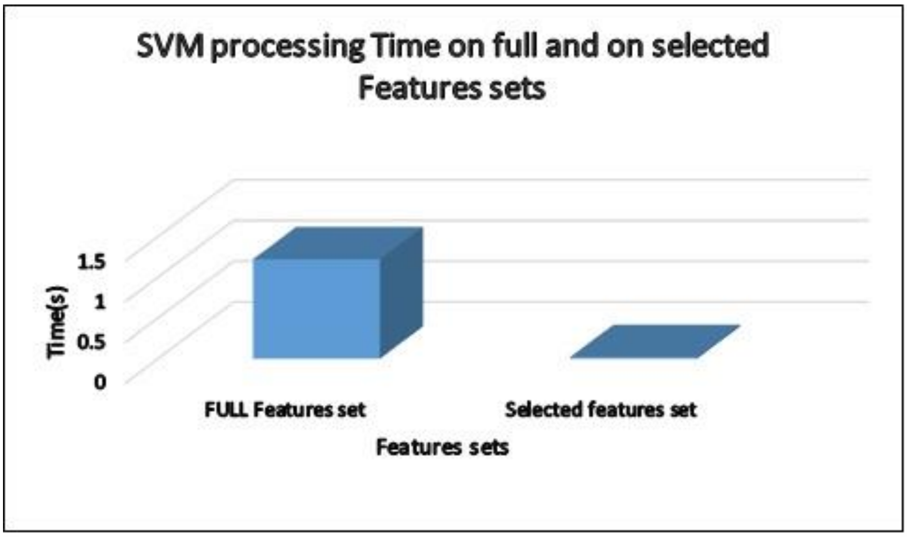

Fig.8 Accuracy of SVM on Full and on selected features sets

3.5. Deep learning model for HD detection

We also use deep learning model for detection of HD using full set of feature. The deep learning model with essential parameters has been trained and tested and reported in Table 5. Deep learning architecture such as DNN1, DNN2, DNN3, DNN4 have been shown in Table 5. According to Table 5 DNN3 achieved high accuracy 91.10 % as compared DNN1, DNN2, and DNN4.

Thus, according to Table 4 and 5 the classification performance of ML learning model is high as compared to deep learning model. ML learning algorithms need features selection for improving accuracy while deep learning automatically select important features from the data set. The performance of deep learning model on this data set is low because the data set instances are small and deep learning need big data set for deep learning, that why the prediction performance of deep learning on this data set is low. The performance of deep learning model shown in figure 9.

Table 5 Deep learning model for classification of HD

Network	Training instance	Testing instance	Learning rate	Activation function	Epochs	Timing of Training(s)	Prediction accuracy(%)
DNN1	210	87	0.0001	RELU	300	120	89.13
DNN2	210	87	0.0010	RELU	600	130	90.12
DNN3	210	87	0.0011	RELU	800	150	91.10
DNN4	210	87	0.0010	RELU	900	170	90.70

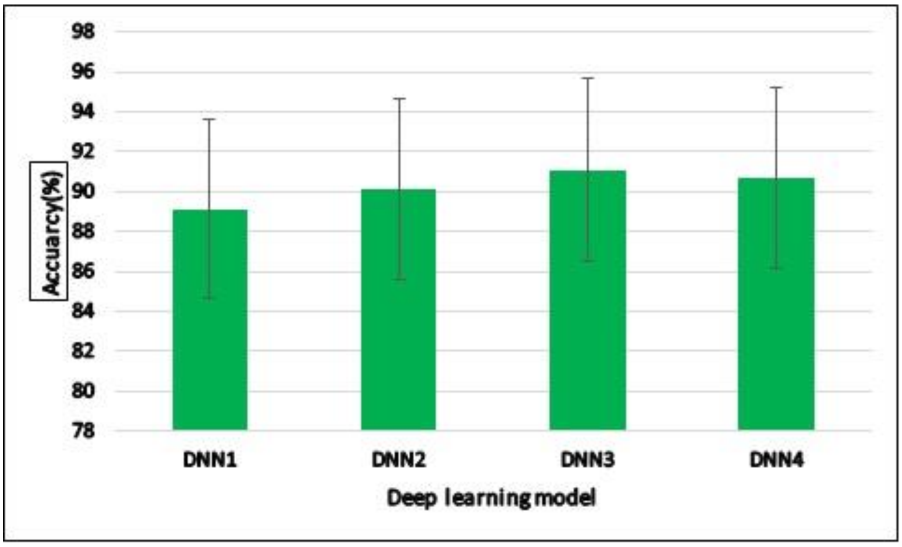

Fig.9 Deep learning performance

3.6. Proposed method performance comparison with existing state of the art methods

The performance proposed method (Chi-squared-SVM) have been compared with state of the art method in Table 6. According to Table 6 the proposed method of HD diagnosis performance was good as compared to existing methods of HD. The Figure 10 show the comparison of accuracy of the proposed method and existing diagnosis methods of HD.

Table 6 Accuracy comparison of proposed method and state of the art methods

Ref	Method	Accuracy(%)
[17]	ANN	88.12
[15]	MLP and SVM	80.41
[32]	Hybrid method	92
[33]	SBF-K-NN	90
Proposed method	Chi-square-SVM	93

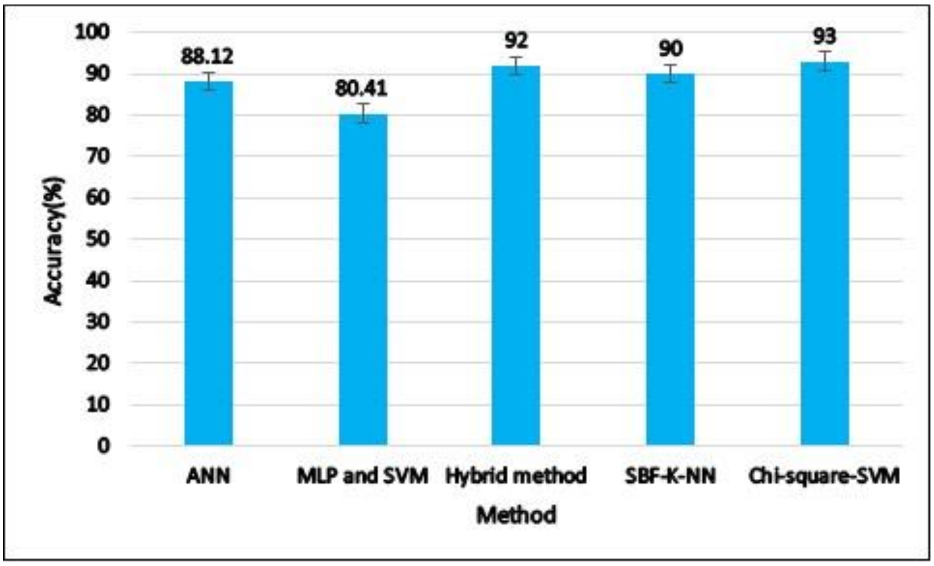

Fig.10 Comparison of accuracy of proposed method and state of the art methods

4. Conclusion

In this research study we proposed heart disease recognition method (Chi-Square-SVM) for efficient detection of heart disease. In the proposed method machine learning models were used for classification of healthy and heart disease subjected rom cleaved data set. We proposed Chi square algorithm for related features selection to improve the classification performance of the models. The experimental results demonstrated that proposed (Chi-Square-SVM) method performance was excellent in term of accuracy and achieved 93% accuracy. Furthermore, performance of proposed method is high as compared deep learning model. The proposed method performance is good as compared to existing heart disease diagnosis methods. The proposed method can easily have incorporated in health care for heart disease identification.

The novelty of the proposed method described as firstly, Chi square feature selection algorithm proposed for related features selection and the features selected by proposed FS algorithm the classifier SVM achieved high accuracy.

Secondly, the performance of SVM is more high as compared to other ML models, so SVM is suitable classifier for HD detection using Cleveland HD dataset. Lastly, the features have low impact in detection of HD are successfully discarded by proposed FS algorithm.

In future, we will use more feature selection and optimization techniques to select more important features from data set to further improve the prediction accuracy of the HD diagnosis system. The proposed method will also apply on other diseases data sets from their detection, such as breast cancer, Parkinson disease. The proposed technique will also deeply on animal and plant disease detection.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 61370073), the National High Technology Research and Development Program of China (Grant No. 2007AA01Z423), the project of Science and Technology Department of Sichuan Province.

Availability of data and materials

The dataset used in this research work available on UCI machine learning repository.

Competing interests

The authors declare that they have no competing interests

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

References

[1] A. L. Bui, *et al.*, "Epidemiology and risk profile of heart failure " *Nature Reviews Cardiology,* vol. 8, pp. 30-41, 9,November 2011.

[2] P. A. Heidenreich, et al., "Forecasting the Future of Cardiovascular Disease in the United States:A Policy Statement From the American Heart Association," Journals - American Heart Association, pp. 933-944, 1, March 2011.

[3] M. Durairaj and N. Ramasamy, "A Comparison of the Perceptive Approaches for Preprocessing the Data Set for Predicting Fertility Success Rate," International Journal of Control theory and Applications, vol. 9, pp. 256-260, 2016.

[4] B. A. Mourão-Miranda J1, Born C, Hampel H, Stetter M., "Classifying brain states and determining the discriminating activation patterns: Support Vector Machine on functional MRI data.," Elsevier, vol. 28, pp. 980-995, 4,November 2005.

[5] S. Ghwanmeh, et al., "Innovative artificial neural networks-based decision support system for heart diseases diagnosis," Journal of Intelligent Learning Systems and Applications, vol. 5, pp. 176-183, 29, August 2013.

[6] Q. K. Al-Shayea, "Artificial neural networks in medical diagnosis," International Journal of Computer Science Issues, vol. 8, pp. 150-154, March 2011.

[7] J. López-Sendón, "The heart failure epidemic," Medicographia, vol. 33, pp. 363-369, 2011.

[8] V. K and J. Singaraju, "Decision support system for congenital heart disease diagnosis based on signs and symptoms using neural networks," International Journal of Computer Applications vol. 19, pp. 6-12, ,April 2011.

[9] S. Nazir, et al., "Fuzzy logic based decision support system for component security evaluation," Int Arab J Inf Technol, vol. 15, pp. 1-9, 29,Noverner 2015.

[10] S. Nazir, et al., "Birthmark-based software classification using rough sets," Arabian Journal for Science and Engineering, vol. 42, pp. 859-871, 1, December 2017.

[11] A. Methaila, et al., "Early heart disease prediction using data mining techniques " Computer Science

& Information Technology Journal, vol. 24, pp. 53-59, 2014.

[12] O. W. Samuel, et al., "An integrated decision support system based on ANN and Fuzzy_AHP for heart failure risk prediction," Expert systems with applications ElSEIER, vol. 68, pp. 163-172, 13,october 2017.

[13] R. Detrano, "International Application of a New Probability Algorithm for the Diagnosis of Coronary Artery Disease," American Journal of Cardiology, vol. 64, pp. 304-310, 1989.

[14] B. Edmonds, "Using Localised 'Gossip'to Structure Distributed Learning," 2005, pp. 1-12.

[15] M. Gudadhe, et al., "Decision Support System for Heart Disease based on Support Vector Machine and Artificial Neural Network," in Computer and Communication Technology (ICCCT), International Conference 2010, pp. 741-745.

[16] H. Kahramanli and N. Allahverdi, "Design of a hybrid system for the diabetes and heart diseases," Expert systems with applications, vol. 35, pp. 82-89, 2008.

[17] S. Palaniappan and R. Awang, "Intelligent heart disease prediction system using data mining techniques," in Computer Systems and Applications, 2008. AICCSA 2008. IEEE/ACS International Conference on, 2008, pp. 108-115.

[18] X. Wu, et al., "Top 10 algorithms in data mining," Springer, pp. 1-37, 4,December 2007.

[19] F. E. Harrell and J. (auth.), "Ordinal logistic regression," in Regression modeling strategies, ed: Springer, 2015, pp. 311-325.

[20] K. Lawen, et al., "Interpreting parameters in the logistic regression model with random effects", Biometrics, vol. 56, pp. 909-914, September 2000.

[21] Amin Ul Haq Jian Ping Li Muhammad Hammad Memon Shah Nazir and Ruinan Sun , "A Hybrid Intelligent System Framework for the Prediction of Heart Disease Using Machine Learning Algorithms", Hindawi Mobile Information Systems, Volume 2018, Article ID 3860146, 21 pages https://doi.org/10.1155/2018/3860146.

[22] Amin Ul Haq et.al, Comparative Analysis of the Classification Performance of Machine Learning Classifiers and Deep Neural Network Classifier for Prediction of Parkinson Disease, IEEE, 2018 15th International computer Conference on Wavelet Active Media Technology and Information Processing(ICCWAMTIP), IEEE,14-16 Dec 2018.

[23] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," ACM transactions on intelligent systems and technology (TIST), vol. 2, p. 27, 2011.

[24] H.-L. Chen, et al., "A support vector machine classifier with rough set-based feature selection for breast cancer diagnosis," Expert systems with applications, vol. 38, pp. 9014-9022, 2011.

[25] Amin Ul Haq, Jian Ping Li , Muhammad Hammad Memon, Shah Nazir et al. Feature Selection Based on L1- Norm Support Vector Machine and Effective Recognition System for Parkinson's Disease Using Voice Recordings, IEEE access, April 2019.

[26] https://doi.org/10.1155/2019/5176705

[27] H. Liu, R. Setiono. Chi2: Feature selection and discretization of numeric attributes, Proceedings of 7th IEEE International Conference on Tools with Artificial Intelligence, Herndon, VA, USA, USA, 1995, 388–391

[28] N. Friedman, D. Geiger, M. Goldszmidt. Bayesian network classifiers [J]. Machine learning, 1997, 29(2–3):131–163..

[29] Amin ul haq et al. Heart Disease Identification Method Using Machine Learning classification in E-Healthcare ieee access, 2020, 14

[30] Amin Ul Haq, et al. "A novel integrated diagnosis method for breast cancer detection." Journal of Intelligent & Fuzzy Systems Preprint: 1-16

[31] Amin Ul Haq, et al. "Recognition of the parkinson's disease using a hybrid feature selection approach." Journal of Intelligent & Fuzzy Systems Preprint: 1-21.

[32] Amin Ul Haq, et al. "Intelligent Machine Learning Approach for Effective Recognition of Diabetes in E-Healthcare Using Clinical Data." Sensors 20.9 (2020): 2649.

[33] A. U. Haq, J. Li, M. H. Memon, M. Hunain Memon, J. Khan and S. M. Marium, "Heart Disease Prediction System Using Model Of Machine Learning and Sequential Backward Selection Algorithm for Features Selection," 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), Bombay, India, 2019, pp. 1-4, doi: 10.1109/I2CT45611.2019.9033683.