IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 14, No. 2, April 2025, pp. 1192~1200
ISSN: 2252-8938, DOI: 10.11591/ijai.v14.i2.pp1192-1200

A hybrid feature selection with data-driven approach for cardiovascular disease prediction using machine learning

Thoutireddy Shilpa, Rajib Debnath
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Hyderabad, India

Article Info

Article history:

Received Mar 21, 2024
Revised Oct 18, 2024
Accepted Nov 14, 2024

Keywords:

Cardiovascular disease prediction
Clinical decision support system
Feature selection
Hybrid feature selection
Machine learning

ABSTRACT

Affecting various disorders of heart and blood vessels mainly cardiovascular diseases (CVDs) is the leading cause of human mortality on the planet. A number of machine learning (ML) based supervised learning approaches existing in the literature have been found useful in the clinical decision support system (CDSS) for detecting CVDs automatically. The challenge, however, is that their performance tends to decline unless the training data is of a certain standard. Several approaches to solving this problem are known as feature selection techniques. Despite several notable advancements in the CVD modeling literature, a weak compendium of research exists in an area which supports the integration of the feature selection approach as a means of enhancing the training quality and thus the prediction accuracy. Against this background, in this paper, we proposed a framework called the cardiovascular disease prediction framework (CVDPF) that integrates ML methods. To support this, we designed and proposed a new hybrid feature selection (HFS) algorithm that aims to reduce the number of parameters. This algorithm adopts several filter methods in order to enhance its performance for the task of feature selection. To improve the prediction accuracy of CVDs, a number of ML tools using the HFS approach has been designed and is termed as machine learning based cardiovascular disease prediction (ML-CVDP). The validation of the framework and the algorithms discussed has been done on the basis of a CVD dataset. The experimental findings demonstrated that CVDPF in combination with HFS outperforms other methods of feature selection available.

Corresponding Author:

Rajib Debnath
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Hyderabad, Telangana 500075, India
Email: rajibdebnath.cse@gmail.com

1. INTRODUCTION

Kolukisaet *et al.* [1] introduced a combined feature selection methodology in diagnosing risks of coronary heart disease (CHD). Ghosh *et al.* [2] reviewed modelling Lasso and Relief based feature selection methods with machine learning (ML)-based methods in prediction of cardiovascular diseases (CVD). Mohan *et al.* [3] put forth a proposal for feature excludings and assessment of some CVD prediction ML techniques. Nourmohammadi-Khiaraket *et al.* [4] presented advanced techniques concerning feature selection to enhance the diagnostic system for the health care system against heart diseases. Nasarian *et al.* [5] presented a heterogeneous hybrid methodology concerning feature selection and diagnosis of coronary artery disease through ML models. To address the problem of heart disease Long *et al.* [6] created a more efficient approach using firefly optimization techniques. Ahmed *et al.* [7] investigated self-aggregating cultural framing among users of health-focused social media to unearth myocardial infarction. Jain and Singh [8] made use of a variety of feature selection and classification techniques in predicting chronic diseases. Yekkala *et al.* [9] were the first to predict heart disease in an enhanced way using an ensemble approach and particle swarm optimization (PSO) technique. Farooq and Hussain [10] suggested a technique called 'machine learning [11] driven prognostic system (MLDPS)' to treat patients with CVDs. Meenakshi *et al.* [12] applied supervised ML techniques, including support vector machines (SVM) [13], random forest (RF), and naive Bayes, to anticipate heart-related illnesses. Latha and Jeeva [14] examined approaches of ensemble classification methods in terms of risk prediction of cardiovascular diseases. Thejas *et al.* [15] introduced a feature selection method which incorporates filter and wrapper methods using mutual information and mini-batch normalization. Ali *et al.* [16] incorporated modern AI methods [17] to achieve the implementation of automatic diagnosis of heart disease. An optimized RF model was developed in [18] and random search algorithm was applied to obtain an intelligent system for CVD prediction. Islam *et al.* [19] also utilized ML model in constructing a framework for CVD forecasting. Diwakar *et al.* [20] cover the latent approaches and trends of ML regarding the automatic detection of CVDs in view of the internet of medical things (IoMT). Khan and Algarni [21] described a prediction model for heart disease while Abdar *et al.* [22] developed a nested ensemble model to diagnose patients with coronary heart disease. Gupta *et al.* [23] implemented MIFH, a ML based diagnostic framework for heart disease that adopts a data driven strategy [24].

To address the problem of heart failure (HF) dynamics, Ali *et al.* [25] came up with a methodology using a stacked ensemble of SVMs [26]. While developing systems for CVD, Kumar *et al.* [27] turned to ML classifiers. Zarkogianni *et al.* [28] suggested yet another new architecture that integrates self-organizing maps (SOM) with hybrid wavelet neural networks (HWNN) to enhance predictive abilities. Ketu and Mishra [29] dealt with the issue of CVD's prediction based on the use of ML approaches in their study involving the Arrhythmia dataset. In their work, Nahato *et al.* [30] presented a hybrid strategy involving extreme learning machine and fuzzy sets in the mining and analysis of the clinical datasets for the purpose of ascertaining the various diseases related to the heart. Nikan *et al.* [31] also forecasted the risk posed by premised coronary artery atherosclerosis an advanced technique termed ridge expectation maximization imputation (REMI). Uyar and Ilhan [32] created a model that integrated recurrent fuzzy neural networks (RFNNs) and a genetic algorithm (GA) for diagnosis of heart diseases. Kavsaoğlu *et al.* [33] suggested a method based on the use of ML algorithms for evaluation of hemoglobin concentration levels without invasive means. Hajj and Kyriacou [34] applied the principles of ML techniques on photoplethysmogram in measuring the blood pressure. Kumar and Gandhi [35] introduced a heart diseases prediction model based on ML technology integrated with IoT devices. However, literature indicates that feature selection is of utmost importance. It has been noted that existing techniques for feature selection can enhance the performance of classifiers using ML. Nevertheless, there is still a question of enhancing the efficiency in terms of feature selection.

2. PROPOSED METHODOLOGY

This section of the paper proposes a strategy to predict CVD diseases using other ML algorithm techniques. Also, the section describes a suggested feature selection scheme whose objective is to enhance the prediction capability of the ML models. This research is centered on supervised learning and, more importantly, selection of variables to maximize the performance of CVD prediction. The approach guarantees that already trained ML models have been provided with sufficient and good quality of training data which if used enhances the prediction accuracy.

2.1. Frame work

In the light of practical aspects, a novel framework is formulated and realized as cardiovascular disease prediction framework (CVDPF). It encompasses mechanisms and algorithms that enable the estimation of the likelihood of any CVD occurrence without human intervention. It consists of some reusable components with particular workflow to enable realization of CVD prediction process. It is equally significant to reduce the feature space to eliminate substandard performance. In using the ML models designed for CVD prediction, performance enhancements can bemade, hence we recommend a hybrid feature selection (HFS) technique associated with the framework as shown in Figure 1.

One of the most challenging aspects of ML techniques is the risk of underperformance caused by the enormous amount of feature space; therefore, appropriate dimensionality reduction techniques should be applied. In order to give a practical implementation of ML methods in cardiovascular disease prognosis, this paper presents a HFS method which is incorporated in the framework. As shown in Figure 1, the CVDPF is the framework that incorporates supervised ML and feature selection for the purposes of CVD prediction. In the framework, the aforementioned feature selection technique is applied prior to any effort on building and training a classifier. Feature selection is preceded by pre-processing of the provided CVD dataset which assists in preparing the data for both the training and testing processes. The CVD dataset comprises the training and test datasets in an 80:20 ratio. The proposed HFS method is then applied on the training data to identify the features that have a true impact on the prediction of the class label. Once feature selection is completed, several ML models are used as forecasting mechanisms for CVD.

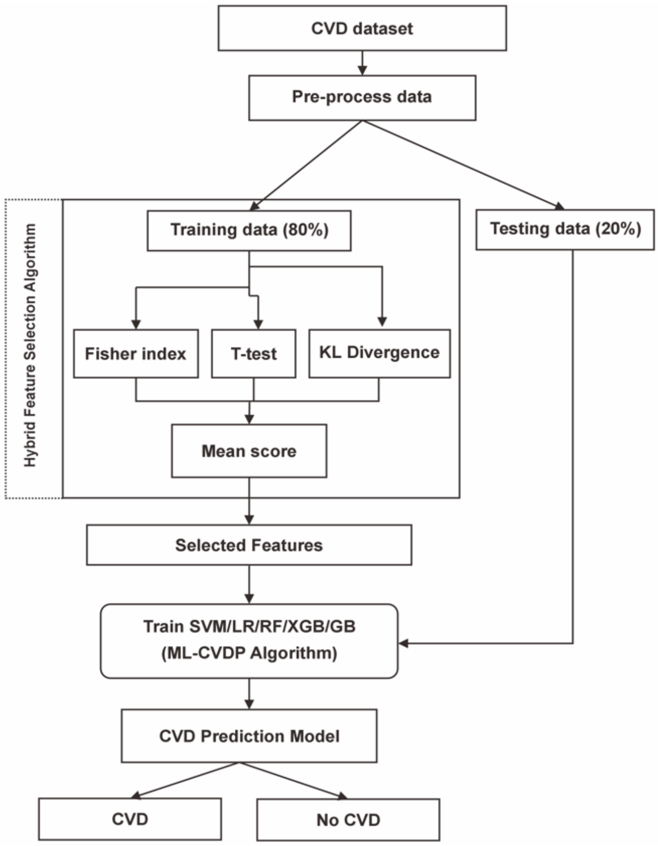

Figure 1. Illustrates overview of the proposed CVDPF

2.2. Dataset details

We have compiled a CVD dataset that comprise essential data about patients, whereby each patient is represented by a single instance. The available features are sufficient for carrying out the diagnosis with a non-invasive procedure. The data set also provides the diagnosis attribute so that the learning process is semi-supervised. Table 1 presents the data set that was used for the empirical evaluation. It has all the necessary features for carrying out an automatic diagnosis of CVD using a ML approach.

Table 1. Details of attributes found in the dataset

Attribute name	What it does?	Value range
ID	Such unique patient identification number.	Unique serial number.
SEX	Contains the gender information of the holders.	Male | Female
AGE	Holds age information	1-100
CP	Holds chest pain type	1-4 (Normal angina, atypical angina, non-anginal pain, asymptomatic)
TRESTBPS	Resting blood pressure value	Some value in mm Hg
CHOL	Serum cholesterol value	Some value in mm/dl
FBS	Fasting blood sugar (greater than 120mg/dl)	This indicates Yes or No by either 1 or 0 respectively.
RESTECG	Holds resting electrocardiographic results	0 or 1 or 2 indicating normal; ST-T wave is abnormal; and probable left-ventricular hypertrophy.
THALCH	Maximum heart rate achieved	Any heart rate value (numeric)
EXANG	Holds exercise induced angina	1 or 0 indicating Yes or No respectively
OLDPEAK	Holds exercise induced ST depression	A numeric value
SLOPE	Holds peak exercise ST segment slope	A numeric value: upsloping/flat/downsloping.
CA	Number of major vessels colored by fluoroscopy	(3-0)
THAL	Holds thalassemia value	7/6/3 reversible defect/ fixed defect/ indicating normal
DIAGNOSIS	CVD diagnosis value	1-4 /0 indicating CVD and No CVD

2.3. Data splitting

The entire collected dataset is subdivided into two parts for the purpose of the supervised learning process. Of the available data, 80% is set aside for training and 20% for testing. In case of testing data, diagnosis attribute is omitted in order to evaluate the predictive capability of the model proposed. The CVD [35] dataset is helpful in the discovery and prediction of whether or not cardiovascular disease is present within the individual.

2.4. Proposed feature selection method

By identifying the relevant features, feature selection helps enhance the quality of the training data. It also reduces number of dimensions hence reducing the learning space which is an advantage in speed and accuracy of prediction. There are quite a number of features selection techniques available such as Lasso and Relief among others. Feature selection from the literature is emphasized. To this effect we offered a method HFS which is a fusion of three filter-based methods. In filter methods importance measuring of the given features should be done. All features taken in as input and the relevant feature subset is identified. After that the features are rated according to some criteria. This cycle is repeated until all features that contribute to the prediction of the class label are determined. The procedure is illustrated in Figure 2.

The HFS that is suggested contains three filter techniques which makes us T-test, fisher criterion and entropy respectively. In the proposed feature selections, less common notation is used which is shown in Table 2. As found in [23], the Fisher criterion is regarded to be one of the best resources for filter feature selection. It is stated in (1).

$$F(i) = \left|\frac{\mu_1(i)-\mu_0(i)}{\sigma_1^2(i)-\sigma_0^2(i)}\right| \tag{1}$$

In a different approach, the t-test that is discussed in [24] is another filter method which can gauge the significance of features, t-test exprestion in (2).

$$t(i) = \left|\frac{\mu_1(i)-\mu_0(i)}{\sqrt{\frac{\sigma_1^2(i)}{n_1}+\frac{\sigma_0^2(i)}{n_0}}}\right| \tag{2}$$

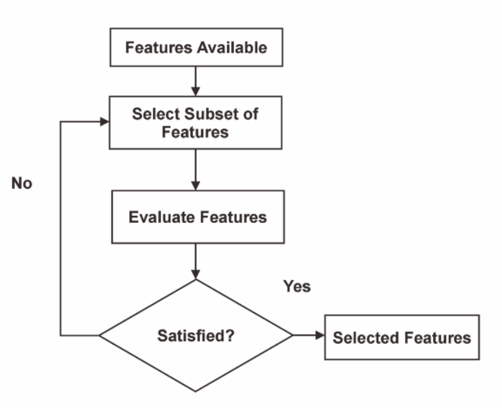

Figure 2. Generic feature selection approach using filter method

Table 2. Notation information

Notation	Description
Q	Probability distribution linked to target
P	Actual probability distribution
KL-distance	Denotes Kullback-Leibler distance
$\mu_1(i)$ and $\sigma_i(i)$	Denotes mean value
n_1 and n_0	Patterns associated with null and unitary class

Another frequent metric that is used as a filter method to evaluate relevant features is referred to as relative entropy. In other words, it is simply the Kullback-Leibler divergence measure already described in [25]. In order to advance a feature selection method, we have come up with a combination of the three metrics. Algorithm 1 also details this method known as HFS. It measures the difference between probability distributions which is represented mathematically in (3).

$$KL(p, q)=\sum_i p_i log_2(\frac{p_i}{q_i}) \quad (3)$$

Algorithm 1: Hybrid feature selection

```
Input: CVD Dataset D, Threshold th
Output: Features Selected F
    a.  Start
    b.  A←ExtractAll(D)
    c.  For each a in A
    d.  f←ExtractFeaturesFromAttribute(a)
    e.  F←F+f
    f.  End For
    g.  For Each f in F
    h.   fscore←ComputeFisherScore(f, F);
    i.  tscore←Compute(f, F);
    j.  escore←ComputeREScore(f, F);
    k.  finalscore←ComputeFScore(fscore, tscore, escore)
    l.  Update M with f and finalscore
    m.  End For
    n.  F←Null
    o.  For each m in M
    p.  IF m.finalscore>=th THEN
    q.  Update F with m.f
    r.  End If
    s.  End For
    t.  Return F
    u.  End
```

2.5. CVD prediction process

Assuming a CVD prediction process, a new algorithm has been designed. This phase is known as ML based cardiovascular disease prediction. It uses HFS algorithm for feature selection and incorporates various other ML models in a sequence for CVD prediction.

Algorithm 2: Machine learning based cardiovascular disease prediction

```
Inputs: dataset D, ml models M
Output: Predictions P
    a.  Start
    b.  (T1, T2)←Preprocess(D)
        Feature Selection
    c.  F←HFSAlgorithm(T1)
        CVD Prediction
    d.  For each model m in M
    e.  Train m with F
    f.  m←FitTheModel(T2, m)
    g.  Add confusion matrix and m to a map C
    h.  Update R with m and predictions
    i.  End For
    j.  For each r in R
    k.  Output prediction results
    l.  Output performance statistics
    m.  End For
    n.  End
```

3. RESULTS AND DISCUSSION

This section will focus on the experimental findings. Among other things, the study consists of CVD prediction observations and performance assessment of the method proposed. The performance of the new method is compared with the feature selection approaches offered previously to static predictive models. The ML prediction models applied are RF, LR, SVM, XGB, and GB. For evaluating the effectiveness of the proposed HFS, various feature selection techniques such as the sequential feature selector, Lasso, RF importance, and Chi-Square are used.

3.1. Precision performance

The use of HFS in different prediction models affects the precision performance positively. It, however, performs differently from other feature selection models when Integrated with the ML models. In consideration to Figure 3, the comparative precision CVD forecasting prediction of different ML methods is displayed when different feature selection strategies are applied. CVD prediction is achieved using ML along with feature selection methods. A comparison of various models in terms of their precision is conducted when different feature selection techniques are applied.

When feature selection is applied, there is a marked improvement in performance. All the feature selection techniques gave better precision as compared to models without such an approach. Furthermore, prediction models that incorporate HFS tend to perform better. The RF model excluding feature selection was precision at 76.2%. The precision metrics for RF are said to be improved when it is applied along with the feature selection methods. For instance, it was '79.41% with Chi-Square, 80% with RF importance, 77.77% with Lasso, 91.17% with sequential feature selector and 92.4% with proposed HFS followed by the highest performance achieved in RF system with HFS'.

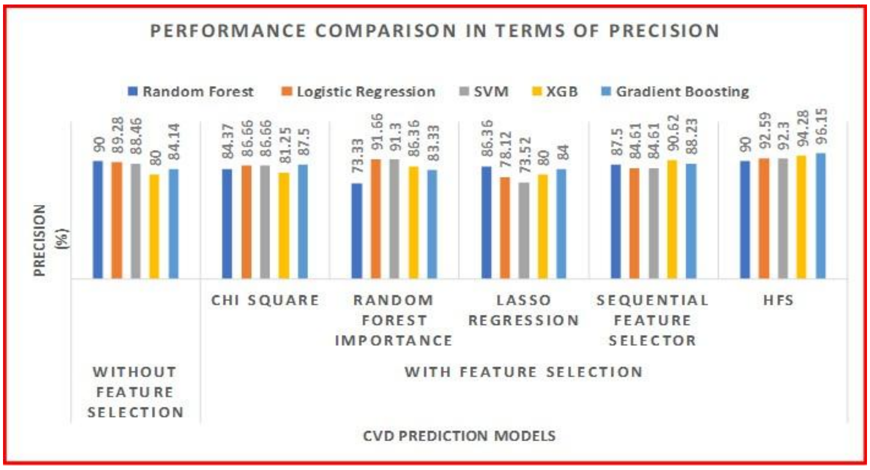

Figure 3. Performance comparison in terms of precision

3.2. Recall performance

The effectiveness of recall comes to light when HFS is used for combination with other models of prediction. Also, it has been shown in the comparison of existing feature selection methods incorporated with the ML models. As can be seen from Figure 4, CVD prognosis applies ML-powered feature selection techniques. The recall of the respective model is estimated in terms of the performance of the applied different feature selection techniques.

Using feature selection results in a considerable enhancement in performance. Every feature selection technique applied enhanced the recall compared to the models that did not employ the use of feature selection techniques. It should be noted that the prediction model's performance utilizing HFS are in most cases higher. The precision achieved by RF without feature selection was 80.28%. Ignoring feature selection while using RF suppresses its performance. For instance, it scored 93.1% recall in case of order Chi-Square, 82.75% in case of sections of RF importance, 95.45% in Lasso, 93.93% in sequential feature selector, and 98.45% in the presented HFS. It is found that the performance of RF with HFS is the maximum.

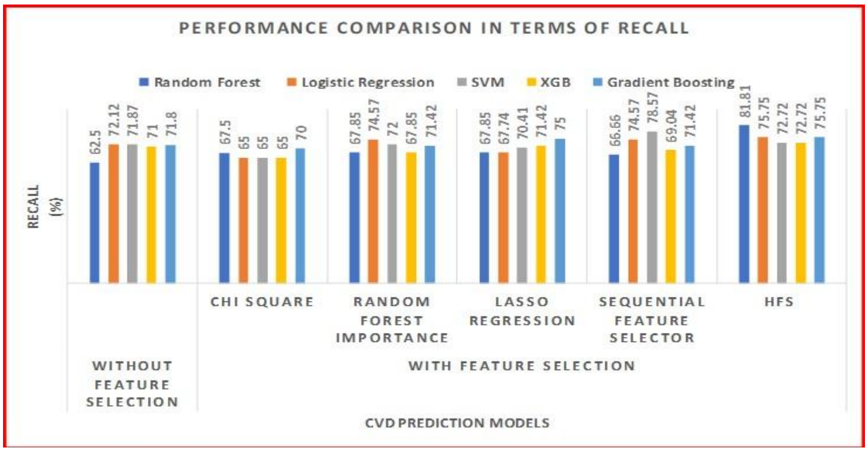

Figure 4. Performance comparison in terms of recall

3.3. F1-score performance

The implementation of HFS in conjunction with different prediction models has been observed to yield an F1-score performance. Its performance is evaluated relative to the available feature selection models used concurrently with ML models. As shown in Figure 5, CVD is predicted using ML and feature selection methods. When various feature selection methods are employed, the F1-score of different models is presented.

The application of feature selection methods leads to noticeable enhancement in performance levels. Every single method of feature selection has recorded a better F1-score than the models that were built without any feature selection at all. Another interesting note is that prediction models HFS perform better comparatively. RF without feature selection had an F1-score of 78.18%. If RF is accompanied by feature selection methods, its performance is enhanced. To begin with, the F1 score recorded by Chi-Square stood at 85.71, whereas that of RF importance was 81.35, by Lasso at 85.7, and by sequential feature selector at 92.52. The proposed HFS scored the highest with a mark of 93.96. Among all these, HFS went best with RF.

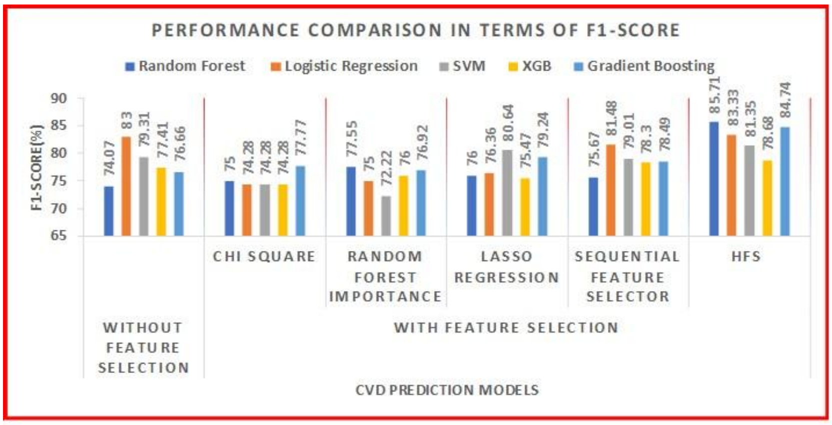

Figure 5. Performance comparison in terms of F1-score

3.4. Accuracy performance

The effectiveness of HFS with various prediction models is noted. Its performance has been evaluated concerning the already established feature selection models that are used with the ML models. In Figure 6, the CVD is predicted with the aid of ML and using Ajala selection methods. There is a comparison in the accuracy of models when different feature selection techniques are applied.

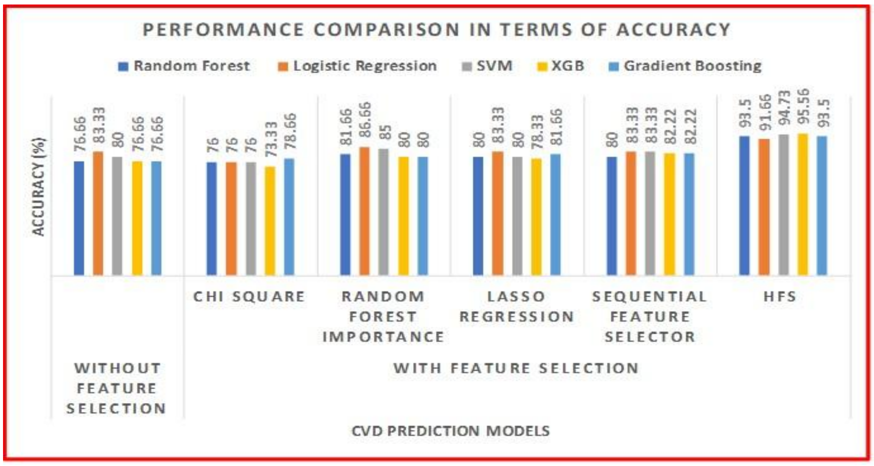

Figure 6. Performance comparison in terms of accuracy

Employing feature selection results in considerable enhancement in performance. Each feature selection technique exhibited improved F1-score as compared to the models which did not incorporate feature selection. One more thing that can be mentioned is that the prediction models which include HFS outperform the rest in terms of performance. The accuracy without feature selection using RF model was 77.79%. RF model performance improves when feature selection methods are applied. The best F1-scores were achieved 85.28% for Chi-Square, 80.94% for RF importance, 85.27% for Lasso, 92.06% for sequential feature selector, and 93.49% for the proposed HFS. The highest value of the performance evaluation metric with HFS is seen in the case of RF.

4. CONCLUSION AND FUTURE WORK

In this section we introduced a system entitled CVDPF over useful ML techniques. Besides, we have also provided a HFS algorithm which is useful in minimizing the effects of the dimensionality problem. This specific algorithm incorporates a number of filter approaches to achieve optimal performance in the process of features selection. Furthermore, we propose an alternative to ML-CVDP or the machine learning-based cardiovascular disease prediction feature. Such a feature will give diverse other machine learning systems and increase the prediction of CVD using the HFS algorithm. Various ML methods, illustrated in the monograph, include SVM, RF, LR, XGB, and GB. Every model thoroughly scrutinizes application of most currently existing feature selection methods along with the newly HFS. The models when used imputing any feature selection algorithm are seen to work efficiently due to the factor of training the process more. A CVD dataset is utilized to test out the effectiveness of the designed structure and methods. The empirical data demonstrated that using HFS with CVDPF is superior to feature selection methods that have been developed for use. Up to now, we have suggested increasing the prediction ability of CVDs employing cabinet approaches for the most effective predictors. A further avenue for future work might also involve the use of medical images and deep learning architectures to predict CVDs.

REFERENCES

[1] B. Kolukisa *et al.*, "Evaluation of classification algorithms, linear discriminant analysis and a new hybrid feature selection methodology for the diagnosis of coronary artery disease," *2018 IEEE International Conference on Big Data, Big Data 2018*, pp. 2232–2238, 2018, doi: 10.1109/BigData.2018.8622609.

[2] P. Ghosh *et al.*, "Efficient prediction of cardiovascular disease using machine learning algorithms with relief and lasso feature selection techniques," *IEEE Access*, vol. 9, pp. 19304–19326, 2021, doi: 10.1109/ACCESS.2021.3053759.

[3] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," *IEEE Access*, vol. 7, no. 04, pp. 81542–81554, 2019, doi: 10.1109/ACCESS.2019.2923707.

[4] J. N. Khiarak, M. R. F. Derakhshi, K. Behrouzi, S. Mazaheri, Y. Z. Harghalani, and R. M. Tayebi, "New hybrid method for heart disease diagnosis utilizing optimization algorithm in feature selection," *Health and Technology*, vol. 10, no. 3, pp. 667–678, 2020, doi: 10.1007/s12553-019-00396-3.

[5] E. Nasarian *et al.*, "Association between work-related features and coronary artery disease: a heterogeneous hybrid feature selection integrated with balancing approach," *Pattern Recognition Letters*, vol. 133, pp. 33–40, 2020, doi: 10.1016/j.patrec.2020.02.010.

[6] N. C. Long, P. Meesad, and H. Unger, "A highly accurate firefly based algorithm for heart disease prediction," *Expert Systems with Applications*, vol. 42, no. 21, pp. 8221–8231, 2015, doi: 10.1016/j.eswa.2015.06.024.

[7] H. Ahmed, E. M. G. Younis, A. Hendawi, and A. A. Ali, "Heart disease identification from patients' social posts, machine learning solution on spark," *Future Generation Computer Systems*, vol. 111, pp. 714–722, 2020, doi: 10.1016/j.future.2019.09.056.

[8] D. Jain and V. Singh, "Feature selection and classification systems for chronic disease prediction: a review," *Egyptian Informatics Journal*, vol. 19, no. 3, pp. 179–189, 2018, doi: 10.1016/j.eij.2018.03.002.

[9] I. Yekkala, S. Dixit, and M. A. Jabbar, "Prediction of heart disease using ensemble learning and particle swarm optimization," in *2017 International Conference On Smart Technology for Smart Nation, SmartTechCon 2017*, 2018, doi: 10.1109/SmartTechCon.2017.8358460.

[10] K. Farooq and A. Hussain, "A novel ontology and machine learning driven hybrid cardiovascular clinical prognosis as a complex adaptive clinical system," *Complex Adaptive Systems Modeling*, vol. 4, no. 1, 2016, doi: 10.1186/s40294-016-0023-x.

[11] M. S. Kumar, H. Azath, A. K. Velmurugan, K. Padmanaban, and M. Subbiah, "Prediction of alzheimer's disease using hybrid machine learning technique," *AIP Conference Proceedings*, vol. 2523, 2023, doi: 10.1063/5.0110283.

[12] B. Meenakshi, B. Gopi, L. Ramalingam, A. Vanathi, S. Sangeetha, and S. Murugan, "Wireless sensor networks for disaster management and emergency response using svm classifier," in *2023 2nd International Conference on Smart Technologies for Smart Nation, SmartTechCon 2023*, 2023, pp. 647–651, doi: 10.1109/SmartTechCon57526.2023.10391435.

[13] R. J. P. Princy, S. Parthasarathy, P. S. H. Jose, A. R. Lakshminarayanan, and S. Jeganathan, "Prediction of cardiac disease using supervised machine learning algorithms," in *Proceedings of the International Conference on Intelligent Computing and Control Systems, ICICCS 2020*, 2020, pp. 570–575, doi: 10.1109/ICICCS48265.2020.9121169.

[14] C. B. C. Latha and S. C. Jeeva, "Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques," *Informatics in Medicine Unlocked*, vol. 16, 2019, doi: 10.1016/j.imu.2019.100203.

[15] G. S. Thejas, S. R. Joshi, S. S. Iyengar, N. R. Sunitha, and P. Badrinath, "Mini-batch normalized mutual information: a hybrid feature selection method," *IEEE Access*, vol. 7, pp. 116875–116885, 2019, doi: 10.1109/ACCESS.2019.2936346.

[16] L. Ali, A. Rahman, A. Khan, M. Zhou, A. Javeed, and J. A. Khan, "An automated diagnostic system for heart disease prediction based on statistical model and optimally configured deep neural network," *IEEE Access*, vol. 7, pp. 34938–34945, 2019, doi: 10.1109/ACCESS.2019.2904800.

[17] R. K. Vanakamamidi, N. Abirami, C. S. Kumar, L. Ramalingam, S. Priyanka, and S. Murugan, "IoT security based on machine learning," in *2023 2nd International Conference on Smart Technologies for Smart Nation, SmartTechCon 2023*, 2023, pp. 683–687, doi: 10.1109/SmartTechCon57526.2023.10391727.

[18] A. Javeed, S. Zhou, L. Yongjian, I. Qasim, A. Noor, and R. Nour, "An intelligent learning system based on random search algorithm and optimized random forest model for improved heart disease detection," *IEEE Access*, vol. 7, pp. 180235–180243, 2019, doi: 10.1109/ACCESS.2019.2952107.

[19] S. Islam, N. Jahan, and M. E. Khatun, "Cardiovascular disease forecast using machine learning paradigms," in *Proceedings of the 4th International Conference on Computing Methodologies and Communication, ICCMC 2020*, 2020, pp. 487–490, doi: 10.1109/ICCMC48092.2020.ICCMC-00091.

[20] M. Diwakar, A. Tripathi, K. Joshi, M. Memoria, P. Singh, and N. Kumar, "Latest trends on heart disease prediction using machine learning and image fusion," *Materials Today: Proceedings*, vol. 37, no. Part 2, pp. 3213–3218, 2020, doi: 10.1016/j.matpr.2020.09.078.

[21] M. A. Khan and F. Algarni, "A healthcare monitoring system for the diagnosis of heart disease in the iomtcloud environment using msso-anfis," *IEEE Access*, vol. 8, pp. 122259–122269, 2020, doi: 10.1109/ACCESS.2020.3006424.

[22] M. Abdar, U. R. Acharya, N. Sarrafzadegan, and V. Makarenkov, "NE-nu-svc: a new nested ensemble clinical decision support system for effective diagnosis of coronary artery disease," *IEEE Access*, vol. 7, pp. 167605–167620, 2019, doi: 10.1109/ACCESS.2019.2953920.

[23] A. Gupta, R. Kumar, H. S. Arora, and B. Raman, "MIFH: a machine intelligence framework for heart disease diagnosis," *IEEE Access*, vol. 8, pp. 14659–14674, 2020, doi: 10.1109/ACCESS.2019.2962755.

[24] K. Karthika, S. Dhanalakshmi, S. M. Murthy, N. Mishra, S. Sasikala, and S. Murugan, "Raspberry pi-enabled wearable sensors for personal health tracking and analysis," in *International Conference on Self Sustainable Artificial Intelligence Systems, ICSSAS 2023*, 2023, pp. 1254–1259, doi: 10.1109/ICSSAS57918.2023.10331909.

[25] L. Ali *et al.*, "An optimized stacked support vector machines based expert system for the effective prediction of heart failure," *IEEE Access*, vol. 7, pp. 54007–54014, 2019, doi: 10.1109/ACCESS.2019.2909969.

[26] B. J. Ganesh, P. Vijayan, V. Vaidehi, S. Murugan, R. Meenakshi, and M. Rajmohan, "SVM-based predictive modeling of drowsiness in hospital staff for occupational safety solution via iot infrastructure," in *2024 2nd International Conference on Computer, Communication and Control, IC4 2024*, 2024, doi: 10.1109/IC457434.2024.10486429.

[27] N. K. Kumar, G. S. Sindhu, D. K. Prashanthi, and A. S. Sulthana, "Analysis and prediction of cardio vascular disease using machine learning classifiers," in *2020 6th International Conference on Advanced Computing and Communication Systems, ICACCS 2020*, 2020, pp. 15–21, doi: 10.1109/ICACCS48705.2020.9074183.

[28] K. Zarkogianni, M. Athanasiou, and A. C. Thanopoulou, "Comparison of machine learning approaches toward assessing the risk of developing cardiovascular disease as a long-term diabetes complication," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 5, pp. 1637–1647, 2018, doi: 10.1109/JBHI.2017.2765639.

[29] S. Ketu and P. K. Mishra, "Empirical analysis of machine learning algorithms on imbalance electrocardiogram based arrhythmia dataset for heart disease detection," *Arabian Journal for Science and Engineering*, vol. 47, no. 2, pp. 1447–1469, 2022, doi: 10.1007/s13369-021-05972-2.

[30] K. B. Nahato, K. H. Nehemiah, and A. Kannan, "Hybrid approach using fuzzy sets and extreme learning machine for classifying clinical datasets," *Informatics in Medicine Unlocked*, vol. 2, pp. 1–11, 2016, doi: 10.1016/j.imu.2016.01.001.

[31] S. Nikan, F. G. Sridhar, and M. Bauer, "Machine learning application to predict the risk of coronary artery atherosclerosis," in *2016 International Conference on Computational Science and Computational Intelligence, CSCI 2016*, 2017, pp. 34–39, doi: 10.1109/CSCI.2016.0014.

[32] K. Uyar and A. Ilhan, "Diagnosis of heart disease using genetic algorithm based trained recurrent fuzzy neural networks," *Procedia Computer Science*, vol. 120, pp. 588–593, 2017, doi: 10.1016/j.procs.2017.11.283.

[33] A. R. Kavsaoğlu, K. Polat, and M. Hariharan, "Non-invasive prediction of hemoglobin level using machine learning techniques with the ppg signal's characteristics features," *Applied Soft Computing Journal*, vol. 37, pp. 983–991, 2015, doi: 10.1016/j.asoc.2015.04.008.

[34] C. El-Hajj and P. A. Kyriacou, "A review of machine learning techniques in photoplethysmography for the non-invasive cuff-less measurement of blood pressure," *Biomedical Signal Processing and Control*, vol. 58, 2020, doi: 10.1016/j.bspc.2020.101870.

[35] P. M. Kumar and U. D. Gandhi, "A novel three-tier internet of things architecture with machine learning algorithm for early detection of heart diseases," *Computers and Electrical Engineering*, vol. 65, pp. 222–235, 2018, doi: 10.1016/j.compeleceng.2017.09.001.

BIOGRAPHIES OF AUTHORS

Thoutireddy Shilpa received B.Tech. degree in computer science and engineering from Vaagdevi College of Engineering, JNTUH, Warangal (D), India, in 2006 and a postgraduate (M.Tech.) degree in computer science and engineering from Amina Institute of Technology, JNTUH, Ranga Reddy (D), India, in 2012. Currently, she is working as Assistant Professor in the Department of B. V. Raju Institute of Technology, Narsapur, Medak, Telangana, 502313, India and Research Scholar in the Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Hyderabad, Telangana, 500075, India. She can be contacted at email: shilpathoutireddy@gmail.com.

Rajib Debnath received Bechelor of Technology (information technology) degree from Bengal Institute of Technology and Management, Shantiniketan, West Bengal, India, in 2010. Masters in Technology (computer science and engineering) and Doctor of Philosophy (Ph.D.) under Department of Computer Science and Engineering at Tripura University (a central university), Suryamaninagar, Tripura, India in 2012 and 2022 respectively. Currently working as an Assistant Professor in Koneru Lakshmaiah Education Foundation (Deemed to be University) Hyderabad, India. His topics of interest are related to the field of computer vision, object detection, machine learning, medical imaging, and image and video processing. He can be contacted at email: rajibdebnath.cse@gmail.com.