RUNDSCHAU

1

A Predictive Model for Heart Disease Risk Using ESVMRF and Advanced Feature Selection Techniques

B. Dhiyanesh¹*, C.B. Selva Lakshmi², T. Yawanikha³, I. Janani⁴, G. Gokila Deepa⁵, K. Chokkanathan⁶

^{1*}Associate Professor, Department of Computer Science Engineering, SRM Institute of Science and Technology, Vadapalani Campus, Tamil, India. Email:dhiyanu87@gmail.com*

²Assistant Professor, Department of Computer Science Engineering, Velammal College of Engineering and Technology, Madurai. Tamil Nadu 625009,India. Email:cbselak08@gmail.com,

³Assistant professor, Department of Information Technology,Karpagam Institute of Technology, Coimbatore, Tamil Nadu641021,India.**Email:yawanikha.it@karpagamtech.ac.in**

⁴Assistant professor, Department of Information Technology, Sona college of Technology, Salem, Tamil Nadu 636005, India. Email:jananimagizh@gmail.com

⁵Assistant professor, Department of Artificial Intelligence and Data Science, PPG Institute of Technology, Coimbatore, Tamil Nadu 641035, India. Email:gokiladeepagopal@gmail.com

⁶Associate Professor, Department of Artificial Intelligence & Data Science Madanapalle Institute of Technology and Science, Madanapalle, Andhra Pradesh 517325, India.Email:chokkanathank2018@gmail.com

Abstract

A major health problem affecting the population worldwide is heart disease. As a result, it is difficult for the heart to function normally and can eventually lead to heart failure, as this condition typically impairs the heart's ability to circulate blood efficiently throughout the body. A significant part of the impact on quality of life is due to the complexity of early heart disease diagnosis methods. A timely diagnosis of heart disease is vital to preventing and treating heart failure effectively. A prognostic method for detecting cardiac abnormalities can be used to determine the likelihood of developing heart disease. However, one of the most critical and complex tasks is to predict heart disease risk factors with sensitivity, specificity, accuracy, and false scores. To solve this problem, we propose an Enhanced Support

RUNDSCHAU

2

Vectorized Machine-based Risk Factor (ESVMRF) technique to more accurately predict heart disease risk levels. Furthermore, we use the Standardized Min-Max Scalar (SMMS) method, a data normalization technique, with pre-processing to analyse the missing attribute values in the heart disease dataset. Furthermore, outliers can be removed by analysing the upper or lower bounds using the Inter-Quartile Range-Based Outlier Removal (IQROR) method. Next, we employ the Relation Feature Weight Vector (RFWV) algorithm, a feature selection technique, to select the optimal features and analyse weights. Using Machine Learning (ML) techniques, the proposed ESVMRF method can be used to predict the risk level of heart patients based on ML techniques to differentiate between healthy patients and patients with heart disease.Further, the method has been shown to be able to predict the risk factors for heart disease patients based on metrics such as sensitivity, specificity, accuracy, precision, F-score, and performance evaluation. The accuracy of this method increased to 95.4%, which provides a valuable and reliable method for predicting heart disease.

Keywords: Heart Disease Prediction, Enhanced Support Vectorized Machine, Risk Factor Analysis, Standardized Min-Max Scalar, Outlier Removal, Relation Feature Weight, Cardiac Abnormalities

1) Introduction

It is impossible to exaggerate the importance of early cardiac illness diagnosis in today's society given the rising incidence of heart disease. It is estimated that 17.9 million people worldwide die each year due to cardiovascular disease, which is the leading cause of death worldwide. Over time, vast amounts of analysed data and hospital patient records have become available. Numerous open-access sources exist for patient records, and various computer techniques are used for analysis to enable accurate patient diagnosis and mortality prevention ¹. Furthermore, efficient data collection and storage are necessary for computer processing. This is a subfield of data mining based on cardiovascular disease datasets that can be efficiently processed utilizing ML techniques. Additionally, ML techniques are employed in the medical industry to diagnose and predict various diseases ².

One important issue that can be resolved with basic physical markers discovered during normal clinical screening is the early prediction of cardiac disease. Accurately improving patient outcomes can be achieved by putting these recommended solutions into practice to decrease the impact of heart disease on people and healthcare systems. Furthermore, these predictive models can identify cardiac patients by analysing extensive clinical data, yielding more precise predictive outcomes than manual diagnosis ³.

¹ Jaswanth Narayana R et al.,, "Heart Disease Prediction System Using Machine Learning", May - 2022.

² Pooja Anbuselvan, "Heart Disease Prediction using Machine Learning Techniques", http://www.ijert.org ISSN: 2278-0181 IJERTV9IS110259.

³ Dengqing Zhang et al., "[Retracted] Heart Disease Prediction Based on the Embedded Feature Selection Method and Deep Neural Network", https://doi.org/10.1155/2021/6260022.

However, predicting the accuracy of high-dimensional data sets is challenging for these systems because there are no proposed approaches to optimize different data features. Early detection of heart disease is crucial for treating patients before a heart attack. Nevertheless, identifying significant risk factors from clinical trials presents challenges for clinicians striving to promptly and accurately diagnose patients ⁴.

The WHO estimates that 17.5 million people worldwide pass away from heart disease and stroke, with low- and middle-income nations accounting for more than 75% of these deaths. Furthermore, stroke accounts for 80% of all heart disease-related deaths, making it a significant contributor to these deaths. Heart illness cannot be properly diagnosed in every instance, and doctors are unable to visit patients around-the-clock due to the increased complexity, time, and understanding required. ⁵.

This paper proposes the ESVMRF method to use these datasets to predict heart disease risk factors more accurately. Furthermore, a heart disease dataset collected from Kaggle can be analysed using ML techniques. The SMMS method can then be implemented to pre-process the heart disease dataset and effectively analyse the missing attribute values. Next, upper and lower bounds can be evaluated, and outliers can be identified and removed using the IQROR technique. Moreover, the RFWV algorithm can select optimal features and analyse their weights. Finally, the proposed ESVMRF method can be used to classify healthy individuals and heart disease patients and accurately predict heart disease risk levels. The potential of this method to improve medical research is inspiring and should motivate further exploration and development.

⁴Ali, F et al., A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion.

⁵Nashif, S., Raihan et al., Heart Disease Detection by Using Machine Learning Algorithms and a Real-Time Cardiovascular Health Monitoring System, doi: 10.4236/wjet.2018.64057.

RUNDSCHAU



Fig 1. The Heart Disease Prediction Basic Architecture Diagram

The basic framework for predicting heart disease is illustrated in figure 1. It involves gathering heart disease data from Kaggle, analyzing it through pre-processing, outlier removal, feature selection, and classification to predict risk factors for heart disease.

2) Literature Survey

According to the study, heart health can be accurately predicted by applying ML algorithms to accurately analyse heart illness. Moreover, the application of ML in public health offers a thorough overview of heart disease prognosis. However, there are a number of difficulties with cardiac diagnosis, including misdiagnosis and treatment delays ⁶. Additionally, ML approaches have the ability to extract features from the Cleveland heart disease dataset in order to estimate the accuracy of heart disease predictions. As a result, the methods described here make substantial use of feature selection and classification model performance ⁷.

⁶Marwah Abdulrazzaq Naser et al.,"A Review of Machine Learning's Role in Cardiovascular Disease Prediction: Recent Advances and Future Challenges," https://doi.org/10.3390/a17020078.

⁷Ahmad AA et al.,Prediction of Heart Disease Based on Machine Learning Using Jellyfish Optimization Algorithm doi: 10.3390/diagnostics13142392.

RUNDSCHAU

6

Furthermore, ⁸ used a DBSCA technique to identify heart problems based on attributes extracted from a dataset used in a medical decision support system. Using ML techniques, a model of a clinical diagnosis assistance system was developed to improve the accuracy of heart disease prediction. The suggested approach aims to automatically derive insights from the clinical data of specific patients. These ⁹findings can improve cardiovascular assessment with the use of UCI repository databases. It is ¹⁰ suggested to test the DNN model with 100 neurons per hidden layer and Relu activation functions in configurations from three to nine layers for early detection of heart disease; it is a massive challenge in its early stages. They also devised a proposed ML-based method to enhance the precision of analyzing heart failure using health parameter information. Moreover, performance could be enhanced by introducing the Principal Component Heart Failure (PCHF) method to identify crucial features in the heart disease dataset ¹¹. The development of large UCI-ML datasets related to heart disease allows for rigorous performance evaluation with powerful models that reliably predict cardiovascular health risks. In addition, it ¹² includes 8,763 samples collected from cardiovascular disease datasets worldwide, providing 26 feature-based numerical and classification variables.

Moreover, the practicality of predicting coronary heart disease risk is reassured by the utilization of two trained DNN. These networks can select different features from the heart disease training dataset, offering a practical and reliable approach ¹³. Additionally, real-world datasets often contain irregular subsets with higher variance than typical data. Nevertheless, creating predictive models for heart disease from such datasets poses a significant challenge. They predict the accuracy of models by using hyperparameter tuning techniques like Randomized SearchCV and Grid SearchCV to select the best performance.

⁸N. L. Fitriyani et al.,"HDPM: An Effective Heart Disease Prediction Model for a Clinical Decision Support System," doi: 10.1109/ACCESS.2020.3010511.

⁹Oumaima Terrada et al.,"Supervised Machine Learning Based Medical Diagnosis Support System for Prediction of Patients with Heart Disease, DOI: 10.25046/aj050533

¹⁰A. A. Almazroi et al.,"A Clinical Decision Support System for Heart Disease Prediction

Using Deep Learning.

¹¹M. Qadri et al., "Effective Feature Engineering Technique for Heart Disease Prediction With Machine Learning, doi: 10.1109/ACCESS.2023.3281484.

¹²Najmu Nissa et al.,"A Technical Comparative Heart Disease Prediction Framework Using Boosting Ensemble Techniques.

¹³T. Amarbayasgalan et al.,doi: 10.1109/ACCESS.2021.3116974.

Extreme Gradient Boosting (EGB) methods are also suggested for determining optimal estimator values based on features ¹⁴. Similarly, the issue can be addressed using mutual information feature selection through the proposal of the LASSO method ¹⁵. Moreover, feature selection approaches can enhance accuracy by decreasing the classification system's processing time.

Heart disease can be predicted during dataset processing utilizing the SVM technique, which can detect correlations between multiple features in the dataset. Furthermore, using a confusion matrix enhances the accuracy of specific, predictable outcomes ¹⁶. Without making these characteristics more complex, a Stochastic Gradient Descent Classifier (SGDC) algorithm is suggested to identify the uncorrelated and imbalanced features in the data ¹⁷. In addition, the method's data assimilation and dimensionality reduction techniques are crucial for reducing expenses and raising model accuracy. They accurately predicted using the Random Forest Bagging Method (RFBM) approach through efficient data collection and optimization of transformation methods on training samples ¹⁸. Nevertheless, accurate predictions may result from a need for more expertise among medical personnel.

¹⁴S. Mondal et al.,"An Efficient Computational Risk Prediction Model of Heart Diseases Based on Dual-Stage Stacked Machine Learning Approaches.

¹⁵ J. P. Li et al., "Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare," .

¹⁶A. Kumar et al., "A Clinical Data Analysis Based Diagnostic Systems for Heart Disease Prediction Using Ensemble Method," doi: 10.26599/BDMA.2022.9020052.

¹⁷ Noor et al., "Heart Disease Prediction Using Stacking Model With Balancing Techniques and Dimensionality Reduction.

¹⁸ P. Ghosh et al., "Efficient Prediction of Cardiovascular Disease Using Machine Learning Algorithms With Relief and LASSO Feature Selection Techniques," doi: 10.1109/ACCESS.2021.3053759.

¹⁹A. Javeed et al., "An Intelligent Learning System Based on Random Search Algorithm and Optimized Random Forest Model for Improved Heart Disease Detection," doi: 10.1109/ACCESS.2019.2952107.

The novel ²⁰ proposed an OCI-DBN that can be optimally configured to increase the performance of heart disease data analysis. Comparisons with other leading methods show that the OCI-DBN model outperforms them in accuracy. An online heart failure database was utilized for feature selection and data testing. A Random Search Algorithm (RSA) was introduced to identify optimal features within constrained diagnostic parameters by employing random forest sampling for predicting heart failure ¹⁹.

Author	Feature	Classificatio	Limitation	Accuracy
	Selection	n		
S.Usha[21]	Recursive Feature	Logistic Regression	Diagnosing heart	77%
	Elimination	(LR), RF	disease can be	
	(RFE)		challenging	
			because there are	
			so many dangerous	
			complications	
Kalaivani K	Auto-Correlation	KNN, SVM	Increases	79.3%
[22]	and Information		healthcare costs	
	gain		and impacts the	
			quality of care.	
Noroozi, Z	Filter Feature	Multivariate Linear	Diagnosing heart	85.5%
[23]	Selection	Model (MLM),	disease is	
	Methods	SVM	challenging due to	
			the numerous risk	
			factors associated	
			with its	
			development.	
Kaushalya	Chi-square,	Gaussian naive	Extracting features	85.71%
Dissanayake	Relief-Feature	Bayes (GNB), LR	from datasets poses	
[24]			a challenging issue	
			when it comes to	
			testing	
Jian Yang [25]	Synthetic	Xtreame Gradient	Identifying	93.44%
	Minority	Boosting (Xgboost)	classification and	
	Oversampling	algorithm	predictive models	
	Technique		for heart disease is	
	(SMOTE)		challenging.	
Gokul Nath,	Genetic	SVM	Obesity and high	88.34%

 Table 1. A Method for Predicting Heart Disease through Feature Selection

 and Classification Based on Machine Learning

9

C.B [26]	Algorithm (GA)		cholesterol are	
			major risk factors	
			for heart disease.	
Hossain MI	C-FSSTBFS	Naïve Bayes (NB),	Serious negative	90%
[27]		K-NN	effects on heart	
			health can even	
			lead to heart	
			failure.	
Anna Karen	CHI-PCA	Random Forests	Symptoms of heart	92.3%
Gárate-		(RF)	disease can be	
Escamila [28]			challenging for	
			doctors to	
			diagnose.	
Cengiz	Correlation-based	NB, SVM	The treatment of	85.14%
Gazeloğlu	Feature Selection		these diseases is	
[29]	(CFS)		expensive.	
Vinta, S. R	GA, LASSO	NB, SVM and RF	Heart disease is a	88.2%
[30]			multifaceted	
			condition that	
			demands	
			significant attention	
			for effective	
			treatment	

The limitations and accuracy were assessed by manipulating the heart disease dataset through ML techniques involving feature selection for cardiac disease prediction, as outlined in Table 1.

²⁰S. A. Ali et al., "An Optimally Configured and Improved Deep Belief Network (OCI-DBN) Approach for Heart Disease Prediction Based on Ruzzo–Tompa and Stacked Genetic Algorithm," doi: 10.1109/ACCESS.2020.2985646.

²¹S.USHA et al., "Predicting Heart Disease Using Feature Selection Techniques Based On Data-Driven Approach," Webology (ISSN: 1735-188X).

²²Kalaivani K et al.,"Heart Disease Prediction System Based on Multiple Feature Selection Algorithm with Ensemble Classifier", DOI 10.1149/10701.8049ecst.

²³Noroozi et al., "Analyzing the impact of feature selection methods on machine learning algorithms for heart disease prediction",https://doi.org/10.1038/s41598-023-49962-w.

²⁴Kaushalya Dissanayake et al., "Comparative Study on Heart Disease Prediction Using Feature Selection Techniques on Classification Algorithms", https://doi.org/10.1155/2021/5581806.

A large medical company uses a Convolutional Neural Network (CNN) algorithm based on Deep Learning (DL) technology to predict mortality in heart failure patients for early detection of heart disease ³¹. However, existing heart prediction systems have several limitations. Huanginitialization and k-mode clustering in the MLP approach ³²are crucial for increasing the precision of heart disease classification. These MLP methods also aid in feature selection from a collection of 70,000 real datasets obtained from Kaggle. The novel [33] proposed K-Nearest Neighbors (K-NN) method tests all features and subsets of a dataset for classification based on feature reduction, evaluation metrics, and processing time. Some techniques are quick but demand significant time for accurate computation³⁴.

The CNN technique is primarily used for cardiac disease prediction and specific tasks to detect complex details in aggregated data ³⁵. However, it is difficult to identify disaggregated areas of heart disease data, which creates challenges of misinformation in publication. Moreover, the precision of cardiac performance techniques can be enhanced by amalgamating heart disease datasets with different classification models while also considering the data monitoring mechanisms and training procedures of diverse mechanisms ³⁶. Next, the Harris Hox Optimization (IHHO) technique was used to select the best features for training on the cardiovascular disease dataset ³⁷. However, the presented methods have some limitations because they require a lot of time for training and testing. Accordingly, a random forest classification algorithm was created to enhance the precision of heart disease identification. The accuracy of training data is approximately 83%, so data analysis is necessary for these applications ³⁸. Subsequently, several classification algorithms have been developed to identify the best combination of features that improve the predictive accuracy of heart disease ³⁹.

²⁵Jian Yang et al.,"A Heart Disease Prediction Model Based on Feature Optimization and Smote-Xgboost Algorithm", https://doi.org/10.3390/info13100475.

²⁶Gokulnath et al.,An optimized feature selection based on genetic approach and support vector machine for heart disease. https://doi.org/10.1007/s10586-018-2416-4.

²⁷Hossain MI et al., "Heart disease prediction using distinct artificial intelligence techniques: performance analysis and comparison," doi: 10.1007/s42044-023-00148-7.

²⁸Anna Karen Gárate-Escamila, et al.,"Classification models for heart disease prediction using feature selection and PCA," Informatics in Medicine Unlocked, https://doi.org/10.1016/j.imu.2020.100330.

²⁹Cengiz Gazeloğlu, "Prediction of heart disease by classifying with feature selection and machine learning methods,"DOI: 10.23751/pn.v22i2.9830.

³⁰Dhiyanesh, B., et al. "Advanced Cloud-Based Prediction Models for Cardiovascular Disease: Integrating Machine Learning and Feature Selection Techniques."

³¹Gopalakrishnan et al., A Novel Deep Learning-Based Heart Disease Prediction System Using Convolutional Neural Networks (CNN) Algorithm. https://ijisae.org/index.php/IJISAE/article/view/3306.

³²Vinta, S. R et al.,Leveraging Machine Learning Techniques for Improving Heart Disease Prediction Systems Using Feature Selection. https://ijisae.org/index.php/IJISAE/article/view/3087.

A. Research Gap

The performance measures offered by the technology available today are inadequate for early diagnosis of heart disease. The performance indicators are improved by comparing the suggested technology to the existing methods. Implementing feature selection can enhance the prediction quality of the approaches mentioned above. Feature selection enhances machine learning models' classification performance by eliminating redundant and noisy features. Feature selection increases the prediction rate, eliminates redundant characteristics, and facilitates the usage of pertinent data for each type of label.

B. Contribution of this work

A Machine learning method framework with feature selection algorithm is proposed and used with a classification algorithm to efficiently analyse and predict heart disease. To leverage prediction accuracy, ESVMRF propose a framework with classifier collections to improving the accuracy. The ESVMRF system predicts cardiovascular disease using a machine learning techniques can manage a wide range of patient data. Developing prediction technologies to increase heart disease risk assessment accuracy and support medical practitioners in making well-informed decisions regarding handling patients. ESVMRF method can be used to predict the risk level of heart patients based on ML techniques to differentiate between healthy patients and patients with heart disease

3) Proposed Methodology

In this work, we greatly improve the accuracy of heart disease risk factors by utilising the capability of the ESVMRF technique for patient classification with heart disease.

³⁹Spencer R et al., "Exploring feature selection and classification methods for predicting heart disease," Doi: 10.1177/2055207620914777.

³³Chintan M et al., "Effective Heart Disease Prediction Using Machine Learning Techniques," https://doi.org/10.3390/a16020088.

³⁴Farhat Ullah et al.,"An Efficient Machine Learning Model Based on Improved Features Selections for Early and Accurate Heart Disease Predication", https://doi.org/10.1155/2022/1906466.

³⁵Fayaz, R et al., An Intelligent Harris Hawks Optimization (IHHO) based Pivotal Decision Tree (PDT) Machine Learning Model for Diabetes Prediction. https://ijisae.org/index.php/IJISAE/article/view/2277.

³⁶Sivakannan Subramani et al.,"Cardiovascular diseases prediction by machine learning incorporation with deep learning," Front. https://doi.org/10.3389/fmed.2023.1150933.

³⁷Sheela et al., Machine Learning based Lung Disease Prediction Using Convolutional Neural Network Algorithm. https://doi.org/10.58496/MJAIH/2024/008.

³⁸Victor Chang et al., "An artificial intelligence model for heart disease detection using machine learning algorithms,"https://doi.org/10.1016/j.health.2022.100016.

Similarly, the RFWV algorithm selects the best features from the dataset and analyses their weights. Finally, the proposed ESVMRF method, with its proven effectiveness, can confidently classify individuals as healthy or heart disease patients and accurately predict their risk levels. The ESVMRF method provides a promising approach to more accurately predict cardiovascular disease risk factors.



Fig 2. The Proposed Architecture Diagram based on ESVMRF

The Kaggle dataset was utilised by the ESVMRF-based architecture diagram depicted in Figure 2 to increase the precision of heart disease prediction. Next, relevant features in the dataset are identified and their relevance is evaluated using the RFWV method. The proposed ESVMRF method can classify individuals as healthy or cardiac patients while accurately predicting their risk levels. This ESVMRF method can predict heart disease risk factors more accurately and classify heart patients.

A. Dataset Collection

2025 123(4)

Data sets related to heart diseases obtained from Kaggle can be analyzed using ML techniques. A selection of 14 features from all published trials can be used to validate the 76

12

features in the heart disease dataset. Furthermore, if a patient has heart disease is indicated by a target heart disease signature. Additionally, integer values can be categorized as 0 = no disease and 1 = disease utilizing the Cleveland database, which is based on ML approaches in these trials. If the patient has cardiac disease, it is indicated via an attribute called "target," which is expressed as an integer between 0 and 4. Understanding the values of "target" attributes offers valuable insights to enhance diagnostic accuracy, treatment approaches, and overall patient care.In previous analyses employing heart disease databases, the main focus was on distinguishing heart disease patients classified as 1, 2, 3, and 4 from patients with no signs of heart disease.

Feature Attribute	Description
Age	Ager in year
СР	Chest Pain
trestbps	Resting Blood Preasure
FBS	Fasting Blood sugar
Chol	Serum Cholesterol in mg/dl
thalach	Achieved maximum heart rate
thal	0=Normal; 1=fixed defect; 2=reversable
	defect

Table2.Dataset	Attribute	Information
----------------	------------------	-------------

As shown in Table 2, the accuracy of these can be predicted by features such as trestbps, FBS, Chol, thalach, and thal using the heart disease dataset

1	age	sex	ср	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	са	thal	target
2	52	1	0	125	212	0	1	168	0	1	2	2	3	0
3	53	1	0	140	203	1	0	155	1	3.1	0	0	3	0
4	70	1	0	145	174	0	1	125	1	2.6	0	0	3	0
5	61	1	0	148	203	0	1	161	0	0	2	1	3	0
6	62	0	0	138	294	1	1	106	0	1.9	1	3	2	0
7	58	0	0	100	248	0	0	122	0	1	1	0	2	1
8	58	1	0	114	318	0	2	140	0	4.4	0	3	1	0
9	55	1	0	160	289	0	0	145	1	0.8	1	1	3	0
10	46	1	0	120	249	0	0	144	0	0.8	2	0	3	0
11	54	1	0	122	286	0	0	116	1	3.2	1	2	2	0
12	71	0	0	112	149	0	1	125	0	1.6	1	0	2	1
13	43	0	0	132	341	1	0	136	1	3	1	0	3	0
14	34	0	1	118	210	0	1	192	0	0.7	2	0	2	1
15	51	1	0	140	298	0	1	122	1	4.2	1	3	3	0
16	52	1	0	128	204	1	1	156	1	1	1	0	0	0
17	34	0	1	118	210	0	1	192	0	0.7	2	0	2	1
18	51	0	2	140	308	0	0	142	0	1.5	2	1	2	1
19	54	1	0	124	266	0	0	109	1	2.2	1	1	3	0
20	50	0	1	120	244	0	1	162	0	1.1	2	0	2	1
21	58	1	2	140	211	1	0	165	0	0	2	0	2	1
22	60	1	2	140	185	0	0	155	0	3	1	0	2	0

Fig 3. Dataset Feature collection

Moreover, as shown in figure 3, the dataset feature collection can detect heart disease using various age, resting, exacting, and old peak features.

B. Standardized Min-Max Scalar (SMMS)

RUNDSCHAU

14

Among the common methods of the standard min-max scalar method, a preprocessing procedure can be done by analysing the missing values of the attributes in the dataset of heart disease. Pre-processing can be used to analyse a dataset on heart disease and accurately determine the data quality. Standard scaler and min-max scaler can pre-process features in the dataset by offering SMMS techniques to handle missing values. Data pre-processing SMMS techniques can assist in identifying missing values to generate smoothed datasets. Heart disease datasets can be managed in several ways, including the removal of missing values, the replacement of numbers with feature values, or the replacement of attribute values with their mean values. The procedure for finding missing values in heart disease data involves removing the variance from the mean, dividing the result by the standard deviation, and then calculating the number of missing values. Normalizing and scaling large eigenvalues through min-max scaling involves first calculating the eigenvalues, then normalizing them by removing the variance from the mean and dividing the result by the standard deviation, and finally scaling them using the min-max scaling technique. This process can also enhance heart disease data. Eigenvalues can also be used to normalize the data within the range of minimum and maximum values.

According to equation 1, missing values can be replaced with mean imputation to eliminate noisy or redundant data. Let's assume N-mean, N-total data in heart disease dataset, w_{μ} -presented in row, W-value.

$$N(w) = \frac{\sum_{w=0}^{M} w_{u}}{1}$$

In equation 2, normalization of heart disease data is commonly used in data pre-processing. The Z-score standardization technique, with a zero mean, can transform data values into a specified range from 0 to 1. Let's assume w'_u –normal data, S_D –standard deviation, N(w) –mean value.

$$w'_{u} = \frac{w_{u} - x(w)}{S_{D}(w)} \tag{2}$$

Calculate the mean and standard deviation of the standardized values as specified in equation 3. The attributes in the provided heart disease dataset include age, trestbps, chol, oldpeak, with escalating dimensional values. Consequently, a standard scale is employed to standardize these attributes to a standard scale. Where σ _standard deviation.

$$\sigma = \sqrt{\frac{1}{M} \sum_{u=1}^{M} (w_u - \mu)^2}$$
(3)

Data can be normalized using a min-max scale by scaling large eigenvalues. As shown in equation 4, normalize infinite intervals between minimum and maximum values to detect missing values. Let's assume w_{min} -minimum value, w_{max} -maximum value.

$$w' = \frac{w_{-W_{\min}}}{(4)}$$

Therefore, to $max-w_{min}$ the efficiency of the analysis, missing values can be filled in using the mean of a user-defined heart disease dataset. All the features are normalized to [0, 1] by using the min and max values of each feature in the heart disease dataset.

RUNDSCHAU 2025 123(4)

C. Inter-Quartile Range-Based Outlier Removal (IQROR)

Outliers can be removed by analysing the upper or lower bounds using a median range-based outlier removal method. Furthermore, the IQROR method can identify the mean value of data points in the intermediate range and eliminate intermediate outliers. IQROR can be determined by subtracting the first or lowest Q1 from the third or highest Q3. Then, these methods can be distributed utilizing the IQROR method to handle outliers in the columns of chol and Trestbps features. Moreover, boxplots can analyse heart disease with outliers where both columns deviate from the normal range. Use the IQROR method to examine a data set's upper and lower limits. Also, after removing the outliers, the data set represents a value of one for the presence of heart disease and a value of zero for the absence of disease.

Algorithm 1. IQROR

Input: Detect missing value w

Output: Remove outlier $Ol_{\mathcal{R}e}$

Begin

1. Calculate the interquartile range

 $Iq_P = W(Iq_{uv})_P$ 2. Outliers can be detected and eliminated.

$$US_{P} = S_{3} - S_{1}$$

For each iteration u=1 to K do

Estimate the upper and lower interval range

$$\mathcal{L}_{\mathcal{B}} = S_{1} - 1.5 * US_{P}$$

$$\mathcal{U}_{\mathcal{B}} = S_{3} - 1.5 * US_{P}$$
For each $((u, v) \in \Omega)$ do
Compute $w_{P}^{K} = US_{u}K_{v}^{K}$
End for each
If $(\mathcal{O} > \mathcal{U}_{\mathcal{B}})(\mathcal{O} < \mathcal{L}_{\mathcal{B}})$ then
Iq_{P} + +;
End if
End for each
Detect the outlier removal
 $\mathcal{O}\ell_{\mathcal{R}e} = \phi;$
For each $((u, v) \in \Omega)$ do
If $(Iq_{uv} \ge \lambda K)$ then
 $\mathcal{O}\ell_{\mathcal{R}e} = \mathcal{O}\ell_{\mathcal{R}e} \cup \{(uv)\}$
End if
End for each

Return $\mathcal{Ol}_{\mathcal{R}e}$ End

As shown in Algorithm 1, outliers can be detected and removed based on the gap between

RUNDSCHAU

upper and lower bound values in the heart disease dataset. Let's assume Iq_P _interquartile range, US_P _interquartile range, w-interquartile training range, $\mathcal{L}_{\mathcal{B}}$ _lower bound, $\mathcal{U}_{\mathcal{B}}$ _upper bound, P-range, $\mathcal{O}\ell_{\mathcal{R}e}$ _outlier removal, K-threshold, u and v-value, λ _outlier identification outlier.

D.Relation Feature Weight Vector (RFWV)

The relation feature weight vector algorithm can analyse feature weights and select the optimal features. Moreover, optimal features can be selected using the RFWV method to estimate their significance to the target variable for binary and multi-class domains. Furthermore, optimal features can be selected, and their weights can be updated to assign weights to each feature in the heart disease dataset. The heart disease dataset can be analysed with large weight values of the optimal features of interest and small weight values of the remaining features. After that, their weights can be defined by closely related classes of the same and different classes. Moreover, RFWV techniques can be utilized to examine the weights of all characteristics by contrasting the values of neighbouring features. The RFWV method can be applied to analyse feature weight vectors and forecast datasets related to heart disease in binary classification.

Algorithm 2. RFWV

Input: Outlier Removal $\mathcal{O}\ell_{\mathcal{R}e}$ Output: Weight vector X Start Compute the amount of training instance $\leftarrow M$ Evaluate the totalnumber of features $\leftarrow e$ Compute the weight vector for the random number of training instances. Select all feature weights $\leftarrow X[0]:= 0.0$ If T: = 1 to m do Select a random target value $\leftarrow p^T$ Assess nearest hit and miss instance $\leftarrow D_N$ If e: = 1 to a do $X[e]:=X[e] - \operatorname{diff}(X, P_T, D)/N + \operatorname{diff}(X, P_T, N)/n$ Evaluate the distinct feature types $\operatorname{diff}(w_e^v, w_f^v) = \begin{cases} 0, w_e^v = w_f^v \\ 1, \text{ otherwise} \\ \operatorname{diff}(w_e^v, w_f^v) = |w_e^v, w_f^v| \end{cases}$

End if

End if

Return Weight vector X End

A weight vector of feature scores can be analyzed by selecting different feature categories and averaging the weights of each feature to calculate the feature quality as shown in

RUNDSCHAU

17

algorithm 2. Let's assume X-weight, M_number instance, a-feature, m-random training instance, D-nearest hit, N-nearest miss, P^T _select random target value, $diff(_{W_e^v, W_f^v})$ _difference type of feature, j-discrete feature.

D. Enhanced Support Vectorized Machine-based Risk Factor (ESVMRF)

An enhanced support vectorized machine-based risk factor method can better identify levels of heart disease risk and classify patients as either heart disease or healthy individuals. The weights of the heart disease dataset can be analyzed by proposing the ESVMRF method, which initializes the sampling distribution by weighting each training sample equally. A further advantage of the ESVMRF is that when dealing with multiple classes, the margin between positives and negatives near the decision hyperplane in Ndimensional space is maximized. Furthermore, the ESVMRF method uses kernel functions to transform data from the input space into a new high-dimensional space. Moreover, determine an optimal decision plane in the kernel function using the ESVMRF method. Furthermore, the ESVMRF technique updates weight distributions and normalizes heart disease datasets. The enhanced features within the selected feature functions can differentiate between heart disease patients and healthy patients, thus providing accurate predictions of the risk status for heart disease patients.

The data normalization and total number of features for the estimated probability distribution are specified in equation 5. Let's assume x_k _probability weight distribution, M-total number of features, u-values.

$$X_{k,u} \leftarrow X_{k,u} / \sum_{u=1}^{M} x_{k,u}$$
 (5)
hown in equation 6, estimate the training error by using a feature to classify the weak

As shown in equation 6, estimate the training error by using a feature to classify the wea learners. Let's assume ξ_k _training error, d_k _train weak classifier feature.

$$\xi_{k} = \sum_{p X_{k,u}} |d_{k} - (w_{u}) - z_{u}|^{2}(6)$$

Calculate the weights based on the performance of the heart disease classification and update and regularize the weight distribution as specified in equations 7 and 8. Where x_k weight, d_k^1 classification performance weight, q_k^1 selected feature weight.

$$\begin{aligned} \mathbf{x}_{k} &= \frac{1}{2} \ln \left[\frac{1}{\xi_{k}^{1}} - 1 \right] \end{aligned} \tag{7} \\ \mathbf{x}_{k+1,u} &\approx \mathbf{x}_{k+u} a^{-\mathbf{x}_{k} \mathbf{z}_{u} \mathbf{d}_{k}^{1} (\mathbf{q}_{k}^{1})} \end{aligned} \tag{8}$$

Compute the feature vector for the selected attributes by analyzing the heart disease dataset equations 9 to 11 describe the optimization of the maximum edge that separates the hyperplane. Where x and g- parameter, w and z- scale parameter feature vector,

$$\begin{aligned} x^{k}w + T &= 0 \qquad (9) \\ & \underset{x}{\min \frac{1}{2}} \begin{pmatrix} x^{k}x \end{pmatrix} \qquad (10) \\ & z_{u} \left(\begin{pmatrix} x^{k}w_{u} \end{pmatrix} + T \right) \geq 1 \end{aligned} \tag{11}$$

Compute the margin values of the positive and negative support vectors as shown in equation

JNDSCHAU

18

11. Let's assume $x^{k}((w_{\perp}) - (w_{\perp}))$ -positive and negative support weight vector,

$$\frac{\mathbf{x}}{|\mathbf{x}||}((\mathbf{W}) - (\mathbf{W}_{-})) = \frac{\mathbf{x}^{k}((\mathbf{W}_{-}) - (\mathbf{W}_{-}))}{\|\mathbf{x}\|} = \frac{2}{\|\mathbf{x}\|}$$
(12)

Compute the optimal plane convex programming problem as depicted in equation 13. Where P-optimal plane, q-support vector, k-iteration.

$$\begin{pmatrix} \min_{x \in P^{r}, \xi \in P+2} \frac{1}{2} (x^{k}x) + g \sum_{u=1}^{M} \xi_{k} \\ q. k. z_{u} ((x^{k}_{W_{u}}) + T) \ge 1 - \xi_{u} \\ \xi_{u} > 0 \end{cases}$$
(13)

The decision of the classifier, as depicted in equation 14, is represented as the summation of boundary support vectors. Let's assume b-feature, S-quadratic program, f-boundary, q-sum of support vector

$$b(w) = q(\sum_{u=1}^{M} z_u \alpha_u s(w_u, w) + f)^{(14)}$$

As shown in equation 15, classify heart disease patients or healthy patients with a linear kernel function. Let's assume $s(W_{11}, W)$ –quadratic linear kernel function, $\varphi(W_{u_1})$ _tranformation function.

 $s(_{W_u, W}) = \phi(_{W_u})\phi(_W)$ (15) A heart disease dataset can be normalized by updating the weight distribution by specifying the sum of the support vectors. In addition, eigenvalues can be used to identify risk factors in heart disease patients by classifying the dataset as heart disease patients or healthy patients.



Fig4. ClassificationESVMRF Method Using Architecture Diagram

The architecture diagram of the suggested ESVMRF technique, which is seen in figure 4, uses the heart disease dataset to compute the feature vectors of biomarkers and predicts accurate risk factors for individuals with heart disease.

Result and Discussion 4)

This section examines the effectiveness of the heart disease classification system and predicts

2025 123(4)

RUNDSCHAU

confusion metrics using the proposed method, which has significant practical implications. In addition, the proposed method offers the possibility of statistical evaluation by calculating heart disease. A comparison with the existing literature on DNN, XGBoost, and SVM methods can be conducted to assess the proposed method's performance. Using the elements of the confusion matrix, one may evaluate the effectiveness of the feature selection technique. Different metrics can be used for evaluating cardiac data, including accuracy, error score, precision, and specificity. The Heart Disease dataset features can be evaluated using the link https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset. The analysis of 1026 records from the heart disease dataset, sourced from the Kaggle repository, was conducted using machine learning techniques, comprising 743 training samples and 283 testing records for heart disease prediction.

Parameter	Value
Name of the dataset	Heart disease
	dataset
Number of Dataset	1026
Tool	Jupyter
Language	Python
Training	743
Testing	283

rapic 5. printiation rarameter	Table 3.	Simulation	Parameter
--------------------------------	----------	------------	-----------

As shown in Table 3, Jupiter notebooks can be created and analysed by implementing the Python programming language while using the simulation parameters for comparative analysis. Additionally, heart disease datasets can be evaluated to identify patients with heart disease and healthy patients.

8						
Heat Disease Pa	Healthy Patient					
Heart disease patient (1)	True Positive (TP)	False Negative (FN)				
Healthy patient (0)	False Positive (FP)	True Negative (TN)				

Fig 5. Confusion Matrix

RUNDSCHAU 2025 123(4)

The TP, FP, TN, and FN values can be analysed to determine the efficiency ratio for the classification problem, as indicated in Figure 5. These elements enable assessing the classification model's effectiveness as it is derived from the confusion matrix.

 $Sensitivity = \frac{Tru^{Posi}}{Tru^{Posi} + Fals^{Nega}} \times 100$ $Precision = \frac{True^{Posi}}{Tru^{Posi} + Fals^{Posit}} \times 100$ $Specificity = \frac{Tru^{Nega}}{Tru^{Nega} + Fals^{Posi}} \times 100$ $False Score = 2 \times \frac{Pre^{cision} * Re^{call}}{Pre^{cision} + Re^{call}}$

 $Accuracy = \frac{True^{Posi} + True^{Nega}}{Tru^{Posi} + Tru^{Nega} + Fals^{Posi} + Fals^{Nega}} \times 100$

Method	Sensitivity	Specificity	Precision	F-	Accura
S				Score	cy
SVM	82.31	82.73	87%	41.23	88.46%
	%	%		%	
DNN	84.56	84.61	90.18	31%	91.39%
	%	%	%		
XGBoos	87.64	89.79	91.23	28%	93.17%
t	%	%	%		
ESVMR	90.28	92.31	93.46	22.18	95.4%
F	%	%	%	%	

2025 123(4)

RUNDSCHAU

Table 4. Heart disease risk factors are predicted using the proposed ESVMRF method and compared with SVM, DNN, and XGBoost. Evaluation based on classification accuracy shows that the ESVMRF method exceeds the other methods. The analysis indicates significant improvements in detecting heart disease and healthy patients, with Sensitivity at 90.28%, Specificity at 92.31%, Precision at 93.46%, F-Score at 22.18%, and overall Accuracy at 95.4%.



Fig 6. Analysis of Sensitivity

RUNDSCHAU

As shown in Figure 6, the accuracy can be predicted by performing the sensitivity analysis and analyzing the heart disease datasets of various selected feature subsets. Furthermore, their accuracy increased to 90.28% in predicting the accuracy of sensitivity analysis using the proposed ESVMRF method. Moreover, comparing the proposed method with the literature-derived methods such as DNN, XGBoost, and SVM, their accuracy improved to 82.31, 84.56, and 87.64%, respectively. Similarly, the techniques presented can enhance the accuracy of sensitivity analysis using heart disease datasets.



73

RUNDSCHAU

As depicted in Figure 7, accuracy prediction involves conducting specificity analysis on the heart disease dataset's various selected feature subsets. Compared to previous methods like XGBoost, SVM, and DNN from existing literature, the accuracy enhancements are significant, reaching 80.73%, 84.61%, and 89.79%, respectively. These techniques showcased here have the potential to enhance specificity in heart dataset analyses. Furthermore, employing the proposed ESVMRF method for accuracy prediction in specificity analysis led to a significant improvement, achieving an accuracy of 92.31%.



In Figure 8, accurate prediction is demonstrated through a detailed analysis of the heart disease dataset across different selected feature subsets. A significant improvement can be achieved by evaluating the accuracy levels achieved by previous methods, such as DNN-87%, SVM-90.18%, and XGBoost-91.23% with the proposed methods. Furthermore, the proposed methods can be used to analyse a heart disease dataset and provide an assessment of improving accuracy. Additionally, utilizing the offered ESVMRF technique for predictive precision analysis resulted in a significant accuracy enhancement of 93.46%.



As described in Figure 9, testing the features on the heart disease dataset involves evaluating the F-scores on subsets of their selected features to provide an accurate prediction. Compared to other methods like SVM, DNN, and XGBoost from existing literature, the proposed method shows lower accuracy scores of 41.23%, 31%, and 28%, respectively. Similarly, employing the proposed ESVMRF method for accuracy prediction in error score analysis results in a 22.18% lower accuracy in emissions classification. These techniques showcased can potentially enhance false scores when assessing heart disease datasets.



Fig.10. Analysis of Accuracy

Figure 10 demonstrates that utilizing features from a heart disease dataset can offer accurate analysis for heart disease patients utilizing selected feature subsets. Additionally, the suggested ESVMRF approach can select heart data features and enhance accuracy to 95.4%. Furthermore, compared to previous methods, the accuracy improves to DNN-88.46%, SVM-91.39%, and XGBoost-93.17%.

5) Conclusion

This paper proposes an enhanced support vectorized machine-based risk factor to predict cardiovascular risk levels more accurately. Moreover, we will use the SMMS method with pre-processing to analyze the attribute missing values in the heart disease dataset. In addition, outliers can be removed by analyzing the upper or lower bounds using the IQROR method. Next, we select optimal features using the RFWV algorithm and analyze feature weights. Lastly, the ML-based ESVMRF approach may effectively identify heart disease patients and healthy individuals based on their risk level. Additionally, by analyzing confusion metrics like sensitivity, specificity, accuracy, precision, false positive score, and performance evaluation, the approach may identify individuals with heart disease risk factors. Furthermore, the proposed ESVMRF method can select the cardiac data features and improve the accuracy to 95.4%. Also, the accuracy increases to DNN-91.39%, SVM-93.17%, and

RUNDSCHAU 2025 123(4)

XGBoost-93.17% compared to the previous methods. The accuracy of the proposed ESVMRF method has increased to 95.4%, offering a valuable and dependable approach for predicting heart disease.