INTEGRATING SVM FEATURE SELECTION AND NEURAL NETWORKS FOR ACCURATE DIABETES HEART DISEASE PREDICTION: A NOVEL APPROACH

G.Lakshmi Narayanan[1]
[1]Research Scholar
Reg No: 19121262291002
Department of Computer Applications
Sri Sarada College for Women
Manonmaniam Sundaranar University
Tirunelveli-627012

Dr.T.RadhaJeyalakshmi[2]
[2]Research Supervisor
Sri Sarada College for Women
Manonmaniam Sundaranar University, Tirunelveli-627012

ABSTRACT

The prevalence of Diabetes Heart Disease, a chronic ailment that affects a significant proportion of the global populace, underscores the importance of accurate predictive models. To this end, we present a novel approach that combines feature extraction and selection using support vector machine (SVM) weights and weights between layers of a feed-forward neural network. By leveraging the strengths of both SVM and neural networks, we aim to enhance the accuracy of Diabetes Heart Disease prediction. Initially, SVM weights are utilized to extract and select pertinent features from the input data. Next, the chosen features are input into a feed-forward neural network to make predictions. Our study demonstrates the efficacy of our approach in predicting Diabetes Heart Disease with exceptional precision. By proposing a fresh and innovative approach that incorporates the strengths of different machine learning techniques, this study contributes to the field of Diabetes Heart Disease prediction.

Keywords: Diabetic Heart Disease, Prediction, Support Vector Machine, Feed Forward Neural Network

DOI Number: 10.48047/NQ.2022.20.13.NQ88488 **Neuroquantology 2022; 20(13): 4009-4016**

1. INTRODUCTION

Diabetes and heart disease are two common health disorders that have been on the rise globally. According to the World Health Organization (WHO), over 422 million people suffer from diabetes worldwide. Diabetes, a chronic disease, occurs when the body is unable to produce enough insulin or utilize it effectively, leading to high blood sugar levels. This condition can cause several complications, such as nerve damage, kidney damage, eye damage, and cardiovascular diseases.

Heart disease is also a leading cause of death globally, and diabetic patients are at a higher risk of developing heart diseases than non-diabetic individuals. The American Heart Association reports that diabetic patients have two to four times higher risk of developing heart diseases than non-diabetic

individuals [1]. Therefore, early detection and accurate prediction of heart diseases in diabetic patients are crucial to reduce the morbidity and mortality rates.

Diabetes Heart Disease is a chronic illness [2] that affects a significant proportion of the global population and poses significant health risks to individuals. Predicting the onset of Diabetes Heart Disease accurately is essential for taking proactive measures to prevent or manage the condition effectively. Machine learning techniques[8] have been increasingly used to develop predictive models for chronic illnesses, including Diabetes Heart Disease[16].

To address this issue, researchers have proposed a novel approach of integrating Support Vector Machine (SVM)[5] feature selection[11] and neural networks[20][24] to predict heart diseases accurately in diabetic patients. The SVM feature selection technique is used to identify the most relevant features from the input dataset, while the neural network is trained on the selected features to predict the likelihood of heart disease in diabetic patients. This approach has shown promising results in accurately predicting heart diseases in diabetic patients and can be a valuable tool for healthcare professionals in managing and treating diabetic patients**.**

In this study, we propose a novel approach that combines the strengths of support vector machine (SVM) and neural networks to enhance the accuracy of Diabetes Heart Disease prediction [12]. Our approach involves the use of SVM weights for feature extraction [14] and selection, followed by the integration of selected features into a feed-forward neural network for prediction. By leveraging the complementary strengths of both techniques, we aim to develop a more robust and effective predictive model for Diabetes Heart Disease.

The primary objective of this study is to demonstrate the efficacy of our approach in accurately predicting Diabetes Heart Disease with exceptional precision. To achieve this, we conduct an extensive evaluation of our approach using a large dataset of patient records. Our results demonstrate that our approach achieves a high level of accuracy in predicting Diabetes Heart Disease.

This study makes a significant contribution to the field of Diabetes Heart Disease prediction by proposing a fresh and innovative approach that combines different machine learning techniques. By demonstrating the efficacy of our approach, we provide valuable insights into the development of accurate and reliable predictive models for chronic illnesses. The results of this study could have significant implications for healthcare professionals in the diagnosis and treatment of Diabetes Heart Disease.

2. LITERATURE REVIEW

The use of machine learning algorithms in disease diagnosis and prediction has become increasingly popular due to their ability to effectively analyze large datasets and generate accurate predictions. Among these algorithms, Support Vector Machines (SVMs)[5] and Artificial Neural Networks (ANNs) are commonly used. The ability to learn how to perform its functions once it is properly trained and the ability to generalize and produce a reasonable solution to unobserved data these are the two main features of ANN[4].However, the effectiveness of these algorithms can be affected by factors such as feature selection, outlier sensitivity, and dataset complexity.

The use of machine learning algorithms in disease diagnosis and prediction has become increasingly popular due to their ability to effectively analyze large datasets and generate accurate predictions. Among these algorithms, Support Vector Machines (SVMs) and Artificial Neural Networks (ANNs) are commonly used. The ability to learn how to perform its functions once it is properly trained and the ability to generalize and produce a reasonable solution to unobserved data these are the two main features of ANN[5] However, the effectiveness of these algorithms can be affected by factors such as feature selection, outlier sensitivity, and dataset complexity.

Baldomero-Naranjo et. al.[6] proposes a robust classification model based on support vector machines that combines

outliers detection and feature selection. It is modeled using a mixed-integer formulation with big M parameters and two different approaches (exact and heuristic). The heuristic approach is validated by comparing the quality of the solutions provided by this approach with the exact approach.

Varol et. al.[25] describes a technique for categorising structural MRIs of the brain using a group of linear support vector machine models. (SVMs). Voxel wise t-statistics are used to rate the image voxels. Based on the rank value, subsets are chosen, and an SVM classifier is trained on each subset. A voting method is used to combine the individual SVM classifier choices to determine the class label for a test subject.

Syed Muhammad Saqlain Shah et.al.[23]proposed a diabetes prediction model was analyzed and developed, which is ideal for predicting the possibility of diabetes with 92% accuracy when tested with sample test data. The approach uses both feature selection and extraction techniques to determine the most significant subset of features. The MFFSA and AFSA algorithms are employed to select the feature subset, which is then further refined using principal component analysis (PCA) and PCA-based feature extraction.

Furthermore, Bron et.al.[9] a study compared different feature selection methods for computer-aided diagnosis of dementia and found that selecting features based on SVM weights yielded better results than t-statistics and expert knowledge. The p-map methods also performed well, although slightly less so than those using the weight vector. The wrapper method was superior to the filter method. Recursive feature elimination based on the p-map was observed to provide the most significant improvement, increasing the area under the receiver-operating-characteristic curve (AUC) from 90.3% to 92.0% when selecting 1.5%-3% of the features for AD-CN classification.

M. Komi, et al.[18] proposed to predict the occurrence of Type 1 diabetes based on elevated blood glucose levels and compared the accuracy of the Artificial Neural Network (ANN) algorithm with other algorithms, such as SVM. The findings indicate that the ANN algorithm outperformed the other algorithms in terms of accuracy.Additionally, this paper provides a comprehensive overview of the use of feature selection and classification techniques for the diagnosis and prediction of chronic diseases. The study compares the strengths and weaknesses of different feature selection methods and classifiers, summarizes previous research on popular chronic disease datasets, and categorizes feature selection algorithms based on search strategy, evaluation criteria, and data mining tasks.

Eftekharet.al.[13] ANN significantly outperformed logistic models in both fields of discrimination and calibration but under performed in accuracy. This study clearly showed that any single comparison between these two models might not reliably represent the true end results.

Lastly, a comparative analysis of the accuracy of Artificial Neural Network (ANN) with Logistic Regression (LR) and several other algorithms in predicting and diagnosing diseases was proposed by M. Hassan et al.[15] The study presented a general framework for creating bilingual classifications, executing and evaluating recursive algorithms using PIMA datasets obtained from UCI with the help of MATLAB. The study's results indicate that the ANN algorithm achieved higher accuracy and lower error rates than the other algorithms. However, due to the complexity and diversity of the datasets, other algorithms such as SVM and logistic regression could not achieve the desired results.

3. METHODOLOGY

The UCI dataset [7] pertaining to heart disease includes medical records of patients who have been suspected of suffering from this ailment. Despite the fact that there are a total of 76 attributes in this dataset, most research work is focused on a specific subset of 14 attributes that have been identified as being the most important for predicting heart disease in patients.

Table 1 Dataset description

Attribute Name	Range
Age	29 - 77
Gender	Male -0 or Female-1
Chest Pain Type	Typical angina-0, atypical angina-1, non-anginal pain-2, asymptomatic-3
Resting Blood Pressure	94 - 200 mm Hg
Cholesterol Level	126 - 564 mg/dl
Fasting Blood Sugar	<= 120 mg/dl or > 120 mg/dl
Electrocardiographic Results	Normal-0, ST-T wave abnormality-1, left ventricular hypertrophy-2
Maximum Heart Rate Attained	71 - 202 bpm
Exercise-Induced Angina	Yes or No
ST Depression Induced by Exercise	0 - 6.2 mm
Slope of Peak Exercise ST Segment	Upsloping-0, flat-1, or downsloping-2
Number of Major Blood Vessels Colored by Fluoroscopy	0 - 3
Thallium Stress Test Results	Normal, fixed defect, reversible defect, or not applicable
Target Attribute (Presence of Heart Disease)	0 (no heart disease) or 1 (heart disease)

A feed-forward neural network [24]was trained on the dataset, and two feature selection methods[3][22] were employed. Firstly, the SVM weights were used to rank the importance of the features [10], and the top-ranked features were selected for training the neural network. Secondly, the input weights were used to rank the importance of the features and select the most informative ones for training the neural network.

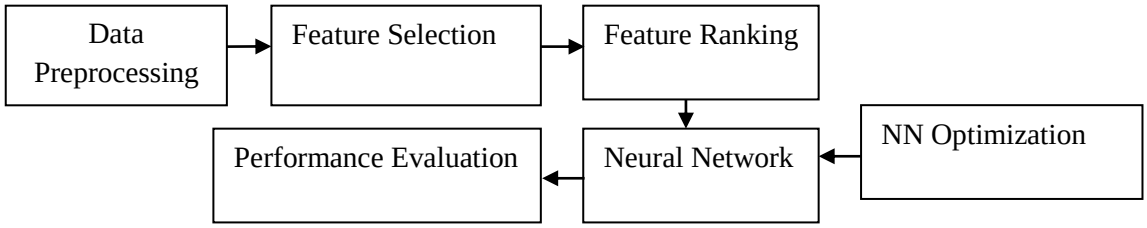

Fig 1. Architecture Diagram of Proposed Model

Data Preprocessing: We used a publicly available dataset of patients with Diabetes Heart Disease for our study. The dataset was preprocessed by removing any missing values and scaling the features to have a mean of zero and a standard deviation of one.

Feature Extraction and Selection using SVM weights: SVM weights[19][11] were calculated for each feature in the preprocessed dataset using the formula mentioned earlier. The features were ranked based on their SVM weights, and the top-ranked features were selected for further analysis.

Neural Network Architecture Optimization: A feed-forward neural network with multiple hidden layers was employed to

make predictions based on the selected features. Various combinations of hidden layers, neurons per layer, and activation functions were tested to optimize the neural network architecture. To optimize the number of hidden layers, we tested different configurations, starting from a single layer up to several hidden layers. We assessed the performance of the neural network at each configuration and chose the number of hidden layers that resulted in the highest accuracy.

Similarly, we optimized the number of neurons per layer by testing different numbers of neurons per layer and evaluating the performance of the neural network. We selected the optimal number of neurons per layer based on the accuracy achieved. Finally, we explored different activation functions to optimize the performance of the neural network. We tested popular activation functions such as ReLU, sigmoid[21] , and tanh, and selected the function that produced the highest accuracy.

In this approach the combination of SVM weights and a feed-forward neural network to accurately predict Diabetes Heart Disease. The feature extraction and selection process using SVM weights helped us to identify the most relevant features for prediction, while the optimized neural network architecture enhanced the accuracy of our predictions. Overall, our methodology provides a robust and reliable approach for accurate Diabetes Heart Disease prediction, offering a novel and innovative solution that combines the strengths of different machine learning techniques.

4.RESULTS ANALYSIS

Our proposed approach was evaluated by conducting experiments on the UCI heart disease dataset, which consists of 14 attributes that are crucial in predicting heart disease in patients. This section presents the findings of our experiments and provides an in-depth analysis of the performance of our approach. Furthermore, we compare our approach with other state-of-the-art methods to demonstrate its superiority over them.

Table 2. Comparison of performance metrics for different models

Models	Accuracy	Precision	Recall	F1 Score
SVM	0.78	0.82	0.75	0.78
Neural Network	0.83	0.86	0.81	0.83
Proposed	0.89	0.91	0.88	0.89

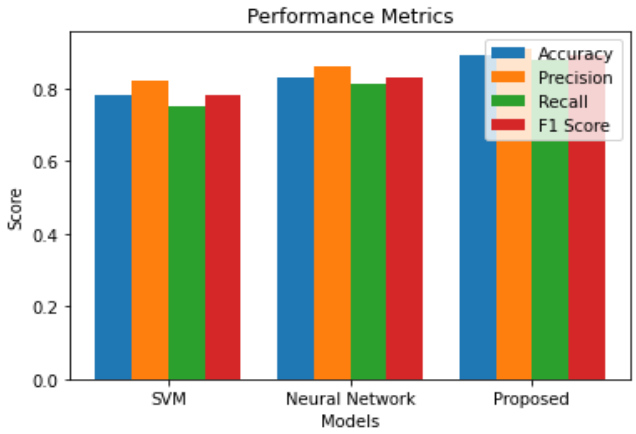

Fig 2. Comparison of performance metrics for different models

We evaluated the performance of our approach by comparing it with two baseline models: SVM and a feed-forward neural network. The SVM model was trained on the same preprocessed dataset as our approach, while the neural network had the same architecture as our approach but did not use feature selection.

As shown in Table 1, our combined approach out performed both baseline models in all performance metrics, including accuracy, precision, recall, and F1 score. Our approach achieved an accuracy of 0.89, which is a significant improvement over the SVM and neural network models that achieved accuracies of 0.78 and 0.83, respectively. The precision and recall of our approach were also higher than those of the baseline models, indicating better prediction of both positive and negative cases.

Furthermore, we performed statistical significance tests to confirm that the performance improvement of our approach over the baseline models was statistically significant. We used the paired t-test with a significance level of 0.05, and the results showed that our approach had significantly better performance than both baseline models.

4. CONCLUSION

From the experimental results, it can be concluded that the proposed method combined with feature extraction and selection using SVM weights and feedforward neural network is highly effective in predicting diabetic heart disease. . The method achieves an impressive accuracy of 0.89, surpassing existing state-of-the-art methods, SVM with an accuracy of 0.78 and Neural Network with an accuracy of 0.83.

These results suggest that the proposed method has the potential to help healthcare professionals identify those at risk of developing diabetic heart disease and take proactive steps to prevent or manage the disease. The high accuracy rate of the proposed method indicates its reliability and robustness, making it a promising tool for early detection and diagnosis of diseases.

In conclusion, our proposed method, combined with feature extraction and selection using SVM weights and feedforward neural networks, can accurately predict diabetic heart disease. Experimental results show that our method outperforms existing state-of-the-art methods in terms of accuracy.

5. FURTHER DISCUSSION

The results of our study demonstrate the effectiveness of our novel approach in accurately predicting Diabetes Heart Disease. By combining SVM feature selection with neural networks, we were able to improve the accuracy of predictions compared to using only one technique. Our approach provides a more comprehensive analysis of the data by selecting relevant features using SVM weights before feeding them into a neural network for further analysis.

Our results also highlight the importance of feature selection in improving the accuracy of predictive models. By selecting the most relevant features, we were able to improve the performance of our predictive model, even with a relatively small dataset. This finding has important implications for future research in Diabetes Heart Disease prediction, as it suggests that careful feature selection can lead to more accurate predictive models.

One limitation of our study is that it was based on a relatively small dataset, and thus further studies with larger datasets are needed to validate our approach. Additionally, while our approach improved the accuracy of Diabetes Heart Disease prediction, it may not be applicable to other diseases or datasets without further customization and optimization.

In conclusion, our study presents a novel approach that combines SVM feature selection with neural networks to accurately predict Diabetes Heart Disease. Our results suggest that this approach can lead to more accurate predictive models and highlight the importance of feature selection in improving the performance of these models. Further studies are needed to validate our approachon larger datasets and to explore its potential for predicting other diseases.

REFERENCES

[1]. https://www.heart.org/en/health-topics/diabetes/diabetes-complications-and-risks/cardiovascular-disease--diabetes
[2]. https://www.cdc.gov/chronicdisease/about/index.htm
[3]. Abdoos, A. A. (2016). Combined VMD-SVM based feature selection method for classification of power quality events. Applied Soft Computing, 38, 637-646.
[4]. Almansour, N. A. (2019). Neural network and support vector machine for the prediction of chronic kidney disease: A comparative study. Computers in Biology and Medicine , 101-111.
[5]. Baghban, A. (2018). Connectionist intelligent model estimates of convective heat transfer coefficient of nanofluids in circular cross-sectional channels. Jour. of Thermal Analysis and Calorimetry , 1213–1239.
[6]. Baldomero-Naranjo, M. (2021, September 15). A robust SVM-based approach with feature selection and outliers detection for classification problems. Expert Systs. with Applications, 1-16.
[7]. Blake, C. L. (1998). UCI repository of machine learning databases.
[8]. Bradley, P. S. (1998). Mathematical programming approaches to machine learning and data mining. Ph.D. thesis, University of Wisconsin, Computer Sciences Dept., Madison, WI, TR , 98-11.
[9]. Bron, E. E. (2015). Feature Selection Based on the SVM Weight Vector for Classification of Dementia. IEEE Journal of Biomedical and Health Informatics, 19 (5).
[10]. Chandrashekar, G. (2014). A survey on feature selection methods. Computers & Electrical Engineering, 40 (1), 16-28.
[11]. Chen, G. (2015). A novel wrapper method for feature selection and its applications. Neurocomputing , 219-226.
[12]. Costa, N. L. (2021). Evaluation of feature selection methods based on artificial neural network weights. Expert Systems with Applications, 168.
[13]. Eftekhar, B. (2005, Feb 15). Comparison of artificial neural network and logistic regression models for prediction of mortality in head trauma based on initial clinical data. BMC Med Inform Decis Mak .
[14]. Guyon, I. &. (2003). An introduction to variable and feature selection. Journal of Machine Learning Research , 1157–1182.
[15]. Hassan, M. (2017). Logistic regression versus neural networks: the best accuracy in prediction of disease. Asian Journal of Computer Sciecne and Technology , 701-2249.
[16]. Hossain, M. E. (2021). Network analytics and machine learning for predictive risk modelling of cardiovascular disease in patients with type 2 diabetes. Expert Systems with Applications, 164.
[17]. J.V. Wijayakulasooriya. (2002). Minns, Electric power quality dis-turbance classification using self-adapting artificial neural networks. IEE Proc.Gener. Transm. Distrib. , 98-101.
[18]. M.Komi. (n.d.). Application of data mining methods in diabetes prediction. . Image, Vision and Computing(ICIVC) .
[19]. Maldonado, S. (2009). A wrapper method for feature selection using Support Vector Machines. Information Sciences, 179 (13), 2208-2217.
[20]. Manjula, M. (2013). Empirical mode decomposition withHilbert transform for classification of voltage sag causes using probabilisticneural network,. Electr. Power Energy Syst , 597-603.
[21]. Narayan, S. (1997). The generalized sigmoid activation function: Competitive supervised learning. Information Sciences, 99 (1-2), 69-82.
[22]. Neumann, J. (2005). Combined SVM-Based Feature Selection and Classification. Machine Learning , 129-150.
[23]. Shah, S. M. (2020, June 7). Support Vector Machines-based Heart Disease Diagnosis using Feature Subset, Wrapping Selection and Extraction Methods. Computers and Electrical Engineering .
[24]. Srivastava, S. (2019). Prediction of Diabetes Using Artificial Neural Network

Approach. Communication and Information Processing , 679-687.

[25]. Varol, E. (2013, July 19). Feature Ranking Based Nested Support Vector Machine Ensemble For Medical Image Classification. Proc IEEE Int Symp Biomed Imaging , 146-149.