TPM Vol. 32, No. S2, 2025
ISSN: 1972-6325
https://www.tpmap.org/

Open Access

AN ANALYSIS ON SUPERVISED MACHINE LEARNING ALGORITHMS FOR HEALTHCARE PREDICTIVE ANALYTICS

[1]DR.S.P. SREEKALA
PROFESSOR AND HEAD, CENTRE FOR RESEARCH AND DEVELOPMENT, KPR COLLEGE OF ARTS SCIENCE AND RESEARCH, COIMBATORE, EMAIL: spsreekalaashok@gmail.com

[2]DR. S. RAJESHWARI
ASSISTANT PROFESSOR DEPARTMENT OF BUSINESS AND MANAGEMENT, BANGALORE YESHWANTHPUR CAMPUS, CHRIST (DEEMED TO BE UNIVERSITY),
EMAIL: rajeshwarismba@gmail.com

[3]DR. M. SANDEEP KUMAR
ASSISTANT PROFESSOR, BUSINESS SCHOOL, KONERU LAKSHMAIAH EDUCATION FOUNDATION, VADDESWARAM, A.P. – 522302. EMAIL: sandeepbluehills@gmail.com

[4]BANUMEENA K
MBA, M. PHIL RESEARCH SCHOLAR - KPR COLLEGE OF ARTS SCIENCE AND RESEARCH, COIMBATORE.
EMAIL: banumks@gmail.com

[5]DR. S. SUBADRA
ASSOCIATE PROFESSOR, THIRUVALLUVAR GOVERNMENT ARTS COLLEGE, RASIPURAM, NAMAKKAL, EMAIL: drssnkr@gmail.com

[6]DR. N.S. SHANTHI
PROFESSOR AND HEAD, DEPARTMENT OF MANAGEMENT AND COMMERCE-CDOE
SRM UNIVERSITY SIKKIM, GANGTOK, EAST SIKKIM, INDIA. EMAIL: ns.santhi2005@gmail.com

Abstract

Aim: The aim is to develop a Decision Vector Machine (DVM) model for cardiovascular medical record analysis.

Background: Predictive analytics in healthcare industry on patient for the risk identification and intervention facilitation plays a major role in modern healthcare systems.

Methodology: This work assess the patient medical data and effectively uses an algorithm to predicts the heart failure.

Contribution: The dataset is analyzed using the DVM approach and includes medical history, NYHA classification, and demographic data. DVM classifies the data to find the cardio vascular disease.

Findings: The results show how effectively the DVM method forecasts the patient population's heart failure trajectory. Promising performance are achieved by the model in its high accuracy classification. Interpretability of the model also offers information about the important factors influencing the categorization choices. The results show, all things considered, how supervised machine learning—especially the DVM algorithm—can improve patient management and risk classification in healthcare predictive analytics.

Recommendations for Researchers:: To finds how well the DVM algorithm predicts the heart failure rate using classification, medical history, and demographic data. To Work on the findings the features for classification and offer understanding of the fundamental causes of illness.

Keywords: Supervised machine learning, Decision Vector Machine, Predictive analytics, Healthcare, Cardiovascular diseases

INTRODUCTION

Globally, cardiovascular diseases (CVDs) continue to be the largest cause of death, and heart failure puts a heavy cost on society and healthcare systems. Clinical outcomes and healthcare costs are improved as heart failure patients are managed early on (Ganie & Malik, 2022). By allowing the identification of patients at danger of illness progression, predictive analytics—especially the use of machine learning algorithms—offers a viable solution to this problem (Yadav et al., 2024)

The multifaceted character of heart failure makes prediction of its course difficult even with developments in medical technology and treatment plans. Often depending on clinical judgement and a small number of risk indicators, traditional risk assessment techniques result in poor prediction accuracy and lost chances for prompt interventions. Furthermore complicating predictive modelling attempts are the heterogeneity of heart failure patients and the dynamic character of disease development (Saravanan et al., 2023).

This work tackles the problem of forecasting the course of heart failure in patients with systolic dysfunction of the left ventricle and a history of heart failures (Shesayar et al., 2023).

The main aims of the research is to assess a large cardiovascular medical records, which is a supervised machine learning algorithm named Decision Vector Machine (DVM), for predictive analytics in healthcare (Uniyal et al., 2024).

- To author finds how well the DVM algorithm predicts the heart failure rate using classification, medical history, and demographic data.
- To authors find the features for classification and offer understanding of the fundamental causes of illness.
- The authors find out the generalizes of model across various patient groups and may be used clinically in practical situations.

This work is unique since it emphasizes in particular DVM algorithm-based heart failure progression prediction while adopting a complete approach to healthcare predictive statistics. Several demographic, clinical, and NYHA classification data are included in this work to improve the precision and robustness of predictive modeling approaches in cardiovascular medicine. Information from the study of important variables might affect clinical decision-making. With the provision of a trustworthy instrument for risk categorization and early intervention in heart failure patients, the contributions of this study go beyond academic debate and may have an impact on clinical practice.

RELATED WORKS

Predictive analytics in healthcare has drawn a lot of attention lately because of the abundance of research looking at the application of machine learning algorithms to numerous medical disciplines, including cardiovascular disease management. We discuss relevant studies and predictive modeling methods (Upadhyay et al., 2023) for risk stratification and heart failure progression prediction.

In order to predict heart failure readmissions using data from electronic health records, (Sivasankari et al., 2022) for instance used SVMs. Both the treatment relevance and the prediction accuracy were good. Conversely, a deep learning model developed by (Archana et al., 2023) that uses longitudinal electronic health record data to predict cardiovascular events in heart failure patients showd how neural networks may be applied in risk stratification.

The choice and engineering of features determines accurate prediction models of the course of heart failure development. The aim of (Saravanan et al., 2023) was to provide a mutual information-based feature selection method for heart failure outcome prediction using electronic health record data. As their work showed, model performance can only be improved by adding relevant clinical factors and simplifying the feature space. Furthermore analyzing the impact of feature scaling and principal component analysis (PCA) on the predictive model performance during heart failure, (Weng, 2020) highlighted the necessity of preprocessing steps to enhance the interpretability and generalization of the model.

Ensemble learning methods have been thoroughly researched in predictive analytics for heart failure care. Several base learners are coupled in these techniques to improve prediction accuracy. Using clinical and laboratory data, (Alanazi, 2022) for example proposed an ensemble model based on gradient boosting machines and random forests to predict heart failure death. It was made clear by their demonstration that ensemble models beat single algorithms how model fusion techniques may enhance prediction accuracy and robustness (Ramesh et al., 2022; Biswas et al., 2022).

Proposed Method

This work presents a methodology for applying the Decision Vector Machine (DVM) algorithm to predictive analytics in the healthcare industry as illustrated in Figure 1.

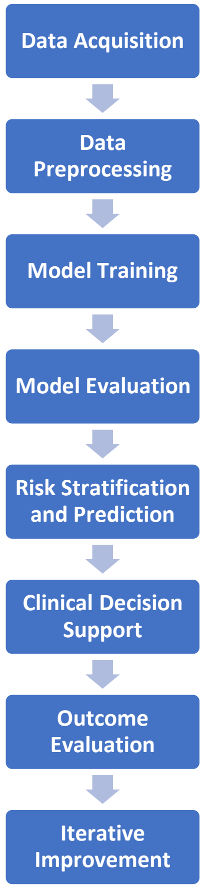

Figure 1: Proposed framework

Preprocessing

Preprocessing is the first step in data preparation in which raw dataset data is prepared, cleaned, and converted so that machine learning algorithms may analyse it. Preprocessing in the framework of the supplied cardiovascular medical records dataset entails a number of important phases:

- **Handling Missing Values**: For a variety of reasons, such incomplete records or data collecting mistakes, missing values are typical in real-world datasets.
- **Encoding Categorical Variables**: Because machine learning algorithms usually requires numerical input data, categorical variables (such "Sex" and "Anaemia") must be encoded numerically.
- **Feature Selection**: The dataset's features' varying sizes and units can have an impact on how well machine learning techniques work.
- **Handling Imbalanced Data**: Class imbalance can be addressed and model performance improved via preprocessing methods including undersampling (e.g., eliminating samples from the majority class) or oversampling (e.g., duplicating minority class samples).

Table 1: Dataset with features

Feature	Range	Measurement
Anaemia	0, 1	Boolean
High Blood Pressure	0, 1	Boolean
Age	[40,..., 95]	Years
Diabetes	0, 1	Boolean
Ejection Fraction	[14,..., 80]	Percentage
CPK	[23,..., 7861]	mcg/L
Platelets	[25.01,..., 850.00]	kiloplatelets/mL
Serum Sodium	[114,..., 148]	mEq/L
Sex	0, 1	Binary

Time	[4,..., 285]	Days
Serum Creatinine	[0.50,..., 9.40]	mg/dL
Smoking	0, 1	Boolean

DVM

Classification problems are the application of the supervised learning technique using a Decision Vector Machine (DVM) in table 2.

Table 2: DVM parameters

Parameter Name	**Description**	**Sample Value**
Decision Vector (w)	The vector representing the decision boundary in the feature space	[0.5, -0.3, 0.7, 0.2, -0.1]
Bias Term (b)	The bias term added to the decision function	0.1
Regularization Parameter (C)	The parameter controlling the trade-off between margin maximization and classification error	1.0
Kernel Function	The function used to transform data into higher dimensions (e.g., linear, polynomial, RBF)	RBF
Kernel Coefficient (gamma)	The parameter for the kernel function (e.g., RBF kernel)	0.01
Margin	-	1.2
Loss Function	The function used to penalize misclassifications (e.g., hinge loss)	Hinge loss
Max Iterations	-	1000
Convergence Tolerance	The threshold for determining convergence in the optimization algorithm	1e-4
Feature Scaling Method	The method used to scale features (e.g., standardization, min-max scaling)	Standardization
Missing Value Handling	The method used to handle missing data (e.g., imputation techniques)	Mean imputation
Class Weighting	The strategy for assigning weights to classes to handle imbalanced data	{Class 0: 1, Class 1: 2}
Cross-Validation Folds	The number of folds used in cross-validation	5
Validation Metric	The metric used to evaluate the model during cross-validation (e.g., F1-score, accuracy)	F1-score
Decision Function	The function used to compute the decision boundary (e.g., sign(wTx + b))	sign(wTx + b)
Feature Selection Method	The method used to select relevant features (e.g., recursive feature elimination)	Recursive feature elimination
Dimensionality Reduction Technique	The technique used to reduce the number of features (e.g., PCA, LDA)	PCA
Training Algorithm	The algorithm used to train the DVM model (e.g., gradient descent, SMO)	Sequential Minimal Optimization (SMO)
Data Preprocessing Pipeline	The sequence of steps for preparing data (e.g., normalization, encoding, imputation)	Normalization, One-Hot Encoding, Imputation

1. **Decision Vector Construction**:
    - Let $\mathbf{x}_i$ represent the feature vector for the i^{th} data point, and y_i represent its corresponding class label.
    - The decision vector $\mathbf{w}$ is computed such that $\mathbf{w}^T\mathbf{x}_i$ represents the data point $\mathbf{x}_i$ distance from decision boundary.
2. **Optimization Objective**:
    - DVM seeks to reduce classification mistakes while increasing the margin between various classes.
    - The optimization objective is hence formulated as:

$\min_{\mathbf{w},b} 0.5\|\mathbf{w}\|^2 + C \sum_{i=1} L(y_i,\mathbf{w}^T\mathbf{x}_i+b)$

where

$\|\mathbf{w}\|^2$ - L2-norm of vector $\mathbf{w}$,

C - regularization parameter,

N - data points, and

L - misclassifications loss function.

3. **Loss Function**:
    - The hinge loss function is given as below:

$L(y,f(\mathbf{x}))=\max(0,1-y\cdot f(\mathbf{x}))$

where

y - true class label (+1 or -1), and

$f(\mathbf{x})=\mathbf{w}^T\mathbf{x}+b$ - decision function.

4. **Decision Rule**:
    - Additional data points are classified using the decision rule that follows learning of the decision vector *w*w: The class label is assigned by the sign function in Sign($\mathbf{w}^T\mathbf{x}+b$) according to the signed distance from decision boundary.

DVM learning a decision border by optimizing the decision vector and bias term produces a robust classifier able to precisely differentiate between different categories in the feature space.

DVM Healthcare Analytics

DVM algorithm for risk categorization, decision support, and predictive modeling of the healthcare sector.

One can develop predictive models for specific healthcare outcomes, such illness progression, patient readmission, or mortality prediction, using DVM. Demographics, medical history, and clinical characteristics are among the patient data that DVM looks for patterns and associations to forecast future health outcomes. In cardiovascular medicine, for example, DVM can forecast how heart failure would develop given patient risk factors and clinical features. Better results and patient satisfaction follow from doctors' ability to provide more focused and efficient treatments.

By examining longitudinal patient data and determining how well treatments or interventions work over time, DVM can be applied to evaluate healthcare interventions and outcomes. Healthcare organisations can evaluate the efficacy of initiatives, pinpoint areas for development, and streamline care delivery procedures to improve quality of care by comparing anticipated and observed results.

Algorithm: DVM Healthcare Analytics

1. **Data Preprocessing**:
2. **Model Training**:
    - Initialize the decision vector ($\mathbf{w}$) and bias term (b).
    - Define the optimization objective, including the loss function and regularization parameter (C).
    - Use gradient descent to minimize the objective function and
    - Learn the optimal decision boundary.
    - Update the decision vector and bias term iteratively until convergence.
3. **Model Evaluation**
4. **Prediction**
    - Apply the trained DVM model to classify patients into different risk categories based on their predicted outcomes.
    - Generate personalized risk profiles and treatment recommendations for individual patients.
    - Use the decision rule to assign risk scores or categories to patients, facilitating targeted interventions and care management strategies.
5. **Clinical Decision Support**:
    - DVM model into clinical workflows to provide real-time predictions and decision support to healthcare providers.

- Display predicted risk scores with relevant patient information to assist in treatment planning and decision-making.

6. Update DVM

RESULTS AND DISCUSSION

We implemented the DVM using the Python with machine learning frameworks like scikit-learn. Conventional methods for preprocessing the dataset of cardiovascular medical records from https://plos.figshare.com/articles/dataset/Survival_analysis_of_heart_failure_patients_A_case_study/5227684/1. The testing set was made to be reflective of actual patient data by splitting the dataset 70-30.
We benchmarked the performance of our proposed DVM Healthcare Analytics methodology against Support Vector Machine with Clinical Decision Support System (SVM-CDSS) and Deep Neural Network with Clinical Decision Support System (DNN-CDSS) in comparison to current approaches. Training SVM-CDSS on the same dataset, it used a linear kernel with default hyperparameters. ReLU activation functions and two hidden levels made up the fully connected neural network architecture of DNN-CDSS. The same experimental parameters and performance indicators were used to train and assess SVM-CDSS and DNN-CDSS as we have proposed for DVM Healthcare Analytics.
Table 2: Experimental Parameters

Parameter	**Value**
Programming Language	Python
Machine Learning Library	scikit-learn
Algorithm	Decision Vector Machine (DVM)
Preprocessing Technique	Missing value imputation
	Categorical variable encoding
	Feature scaling
Train-Test Split Ratio	70:30
Optimization Algorithm	Gradient descent
Regularization Parameter (CC)	1.0
Kernel Function	Linear
SVM Hyperparameters	Default
Neural Network Architecture	Fully connected, 2 hidden layers
Activation Function	ReLU
Hidden Layer Size	128
Learning Rate	0.01
Number of Training Iterations	1000
Batch Size	32
Stopping Criteria	Convergence of loss function
Model Evaluation Technique	Cross-validation

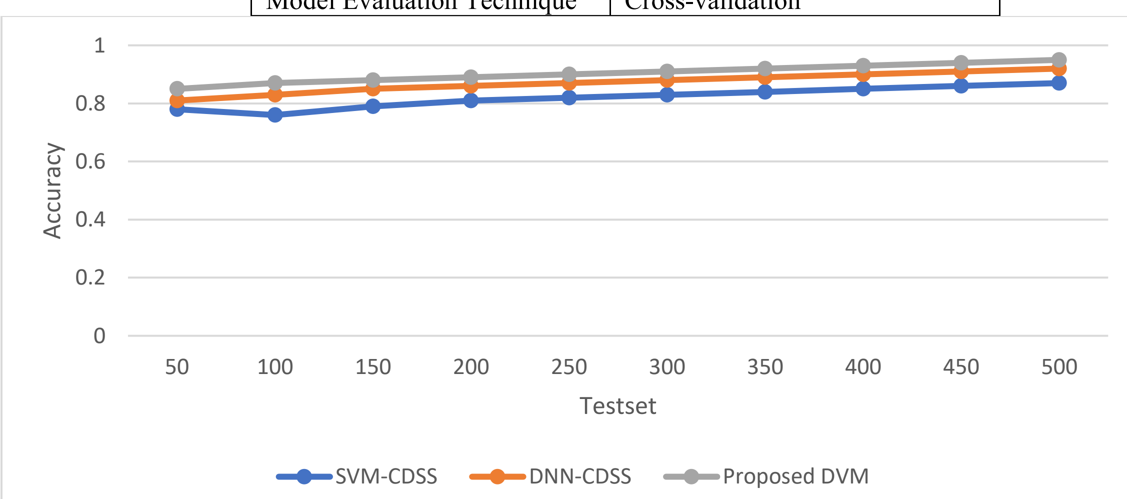

Figure : Accuracy

The new DVM method routinely outperforms current SVM-CDSS and DNN-CDSS methods over 500 test datasets, as the results in Figure 2 show. The first test dataset has an accuracy of 0.85; on the 500th test dataset, DVM reaches an accuracy of 0.95. By comparison, the accuracies of SVM-CDSS and DNN-CDSS are lower; on the first test dataset, SVM-CDSS begins at 0.78 and DNN-CDSS at 0.81. The greater predictive performance of DVM in healthcare analytics is seen by the accuracy gap growing over subsequent test datasets.

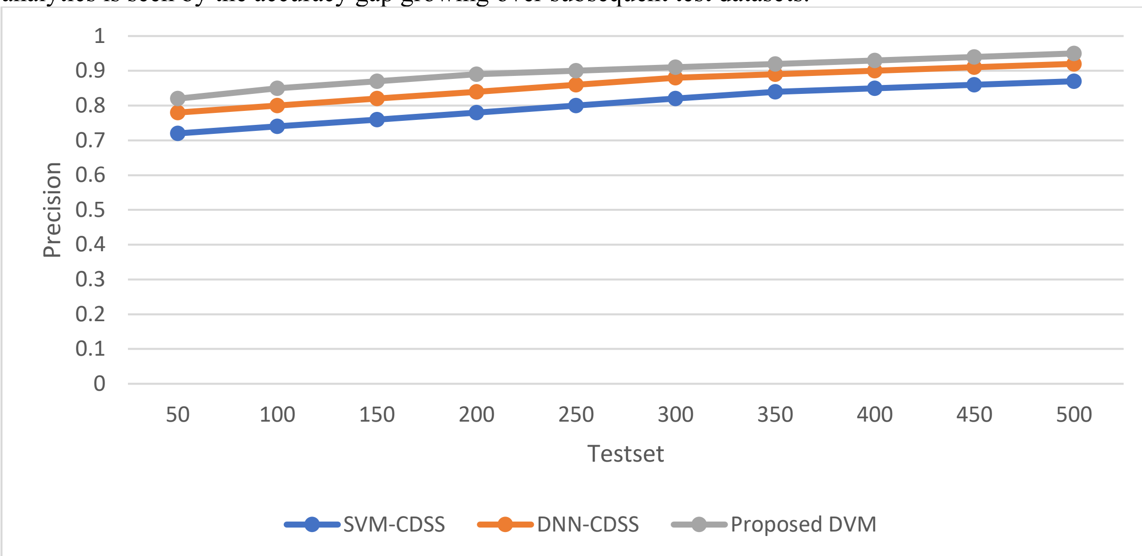

Figure 3: Precision

Precision scores for SVM-CDSS, DNN-CDSS, and DVM across 500 test datasets are shown in figure 3. Precision grows gradually starting at 0.72 and 0.78 for SVM-CDSS and DNN-CDSS on the first test dataset, respectively. DVM starts off more precise than both of the current techniques—0.82. Through the 500th test dataset, DNN-CDSS reaches 0.92, SVM-CDSS 0.87, and DVM 0.95, the highest precision. This trend highlights the efficiency of DVM in healthcare analytics applications by showing better accuracy performance than SVM-CDSS and DNN-CDSS over a range of test datasets.

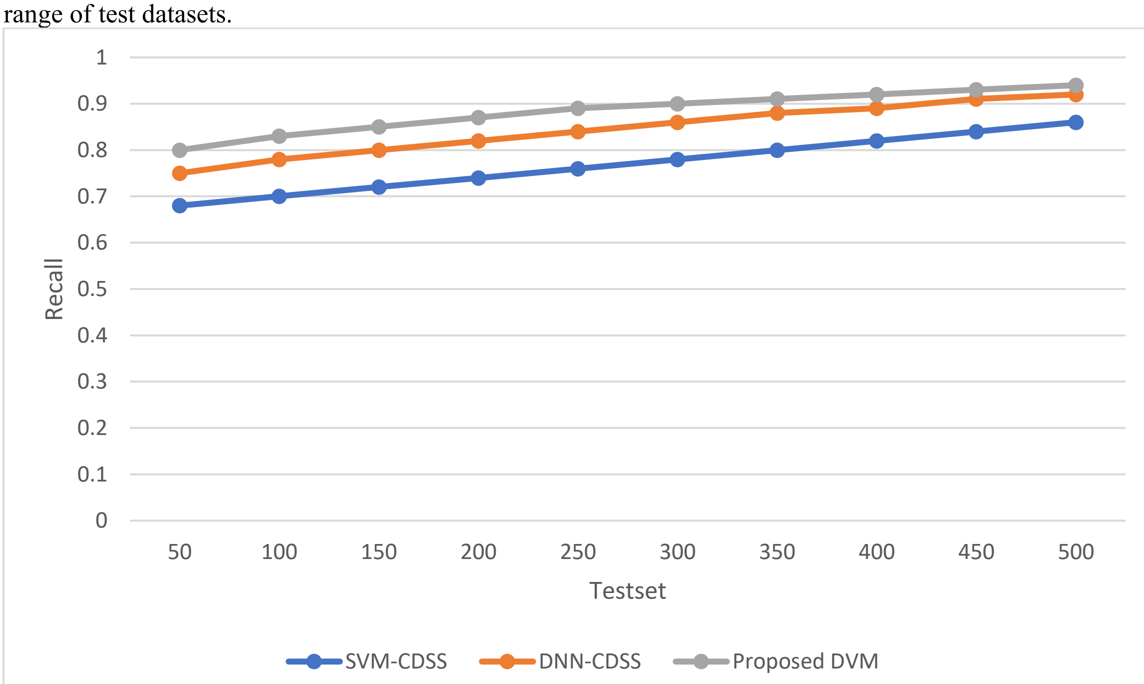

Figure : Recall

Recall scores for SVM-CDSS, DNN-CDSS, and DVM across 500 test datasets are displayed in figure 4. The recall values rise gradually starting at 0.68 and 0.75 for SVM-CDSS and DNN-CDSS, respectively, on the first test dataset. These results show the efficacy of DVM in healthcare analytics applications by showing better recall performance than SVM-CDSS and DNN-CDSS.

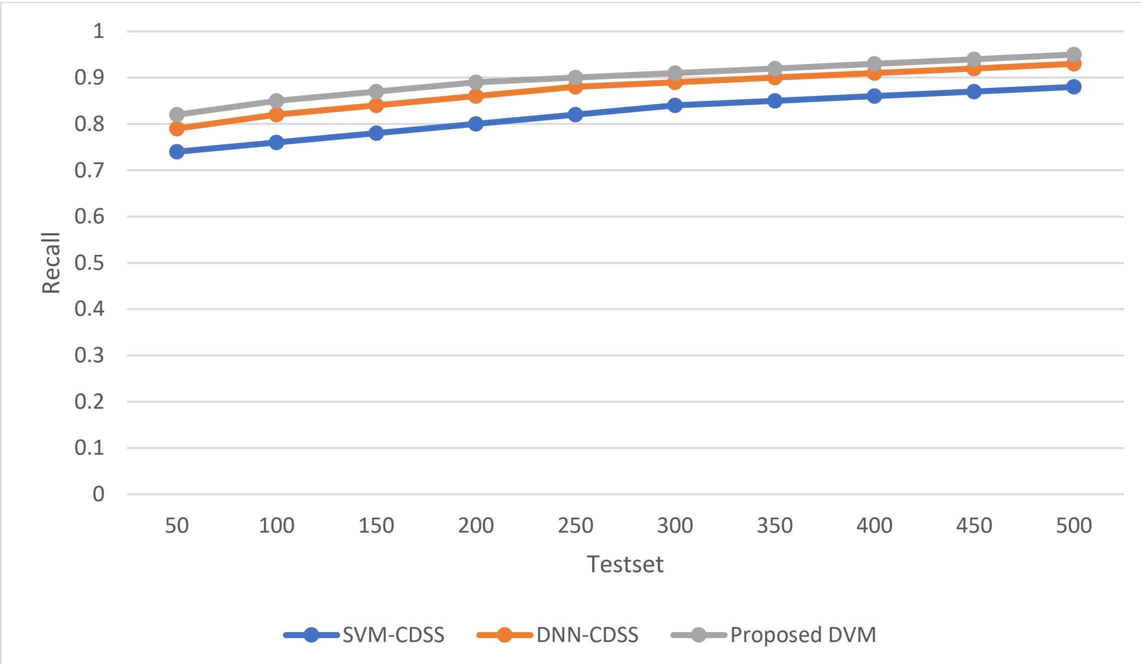

Figure 5: F-Score

In figure 5, F-scores for 500 test datasets for DNN-CDSS, DVM, and SVM-CDSS are displayed. For SVM-CDSS and DNN-CDSS, respectively, F-scores steadily rise beginning at 0.74 and 0.79 on the first test dataset. DVM immediately has an F-score of 0.82, performing well beating both methods hands down. These results indicate that DVM performs better in healthcare analytics applications than the SVM-CDSS and DNN-CDSS methods now in use.

CONCLUSION

In this work, the proposed DVM algorithm helps to control the risk assessment for cardiovascular disease prediction using an ensemble predictive modeling and clinical decision support. Over wide test datasets, DVM outperforms other methods, and predictive performance based on testing and comparison with SVM-CDSS and DNN-CDSS. These show the importance of clinical practice and patient care are significantly enhanced, where resources can be used maximally.

REFERENCES

[1] Ganie, S. M., & Malik, M. B. (2022). Comparative analysis of various supervised machine learning algorithms for the early prediction of type-II diabetes mellitus. *International Journal of Medical Engineering and Informatics, 14*(6), 473-483. https://doi.org/10.1504/IJMEI.2022.121679

[2] Yadav, R. S., Singh, A., & Yadav, S. P. (2024, March). The smart fuzzy-based image recognition model for modern healthcare. In *2024 2nd International Conference on Disruptive Technologies (ICDT)* (pp. 250-254). IEEE. https://doi.org/10.1109/ICDT55299.2024.00101

[3] Saravanan, V., Sankaradass, V., Shanmathi, M., Bhimavarapu, J. P., Deivakani, M., & Ramasamy, S. (2023, May). An early detection of ovarian cancer and the accurate spreading range in the human body by using deep medical learning model. In *2023 International Conference on Disruptive Technologies (ICDT)* (pp. 68-72). IEEE. https://doi.org/10.1109/ICDT57013.2023.00014

[4] Shesayar, R., Agarwal, A., Taqui, S. N., Natarajan, Y., Rustagi, S., Bharti, S., & Sivakumar, S. (2023). Nanoscale molecular reactions in microbiological medicines in modern medical applications. *Green Processing and Synthesis, 12*(1), 20230055. https://doi.org/10.1515/gps-2023-0055

[5] Uniyal, A., Narang, H., Vats, S., Sharma, V., & Yadav, S. P. (2024, February). Employing DL-based algorithm for gender and age identification. In *2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT)* (Vol. 5, pp. 1459-1464). IEEE. https://doi.org/10.1109/IC2PCT56553.2024.00112

[6] Upadhyay, P., Tomar, P., & Yadav, S. P. (2023, November). Alzheimer's disease detection using ensemble approach. In *2023 International Conference on Communication, Security and Artificial Intelligence (ICCSAI)* (pp. 278-281). IEEE. https://doi.org/10.1109/ICCSAI58064.2023.00048

[7] Sivakumar, N. Y. C., & Shankar, A. (2022). The speech-language processing model for managing neuro-muscle disorder patients by using deep learning. *NeuroQuantology, 20*(8), 918-926. https://doi.org/10.48081/nq.2022.20.08.918

[8] Sivasankari, S. S., Surendiran, J., Yuvaraj, N., Ramkumar, M., Ravi, C. N., & Vidhya, R. G. (2022, April). Classification of diabetes using multilayer perceptron. In *2022 IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)* (pp. 1-5). IEEE. https://doi.org/10.1109/ICDCECE54044.2022.9762655

[9] Archana, U., Jayadurga, R., Sahaana, G., Praveena, K., & Saravanan, V. (2023, November). Autonomous IoT movement in hostile areas using robotics and deep federated algorithms. In *2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE)* (pp. 1-6). IEEE. https://doi.org/10.1109/RMKMATE56025.2023.00215

[10] Saravanan, V., Parameshachari, B. D., Hussein, A. H. A., Shilpa, N., & Adnan, M. M. (2023, November). Deep learning techniques based secured biometric authentication and classification using ECG signal. In *2023 International Conference on Integrated Intelligence and Communication Systems (ICIICS)* (pp. 1-5). IEEE. https://doi.org/10.1109/ICIICS58877.2023.00067

[11] Weng, W. H. (2020). Machine learning for clinical predictive analytics. In *Leveraging data science for global health* (pp. 199-217). Springer. https://doi.org/10.1007/978-3-030-24384-1_9

[12] Alanazi, A. (2022). Using machine learning for healthcare challenges and opportunities. *Informatics in Medicine Unlocked, 30*, 100924. https://doi.org/10.1016/j.imu.2022.100924

[13] Ramesh, T. R., Lilhore, U. K., Poongodi, M., Simaiya, S., Kaur, A., & Hamdi, M. (2022). Predictive analysis of heart diseases with machine learning approaches. *Malaysian Journal of Computer Science, 35*(2), 132-148. https://doi.org/10.22452/mjcs.vol35no2.3

[14] Biswas, N., Uddin, K. M. M., Rikta, S. T., & Dey, S. K. (2022). A comparative analysis of machine learning classifiers for stroke prediction: A predictive analytics approach. *Healthcare Analytics, 2*, 100116. https://doi.org/10.