ISSN (ELECTRONIC): 2945-3135

Pramit Kumar Samant, Vinay Pathak & Waqar Ahmad (2026). Machine Learning Based Decision Support System on Electronic Health Record Data to Predict Health Score. International Journal of Multidisciplinary Research & Reviews, 5(7),239-274.

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH & REVIEWS

journal homepage: www.ijmrr.online/index.php/home

MACHINE LEARNING BASED DECISION SUPPORT SYSTEM ON ELECTRONIC HEALTH RECORD DATA TO PREDICT HEALTH SCORE

Pramit Kumar Samant[1], Vinay Pathak[2] & Waqar Ahmad[3]

[1]Department of Computer Science and Engineering, SRMIST, Delhi-NCR, India (pramitsamant@gmail.com)

[2]Department of Computer Science and Engineering, Indian Institute of Information and Technology Sonepat, India (vinaypathak85@gmail.com) **(corresponding author)**

[3]Department of Computer Science and Engineering, Indian Institute of Information and Technology Sonepat, India (waqar.ahmad50@gmail.com)

How to Cite the Article: Pramit Kumar Samant, Vinay Pathak & Waqar Ahmad (2026). Machine Learning Based Decision Support System on Electronic Health Record Data to Predict Health Score. International Journal of Multidisciplinary Research & Reviews, 5(7),239-274.

https://doi.org/10.56815/ijmrr.v5i7.2026.239-274

Keywords

EMR, EHR, Machine Learning, ANN, EDA, SVM, EHR Dataset, Violin plot.

Abstract

Machine Learning as well as AI developed systems are now today greatly used in varieties of real-life applications. E-healthcare management system is also one of them. The decision support system developed using many machine learning algorithms provides to optimize decision in protective health care systems. The proposed system consists of two modules. The first module included a novel Electronic Health Record based decision support system using SVM technology structure. The second module be composed of preprocessing, data extraction, exploratory data analysis from EHR data set. The proposed EHR based decision system is tested with accuracy for detecting health status on 79540 records. EHR data are mostly important for evaluating or predicting the health status after observing the varieties of

	medical parameters like body temperature, heart rate, blood pressure in both range systolic + diastolic and oxygen saturation level etc. How the new sample in the group data analysis can be achieved through machine learning approach and how this approach can be used as in AI tool for E-healthcare system. Also, visualization of data (exploratory data analysis i.e., EDA) is important to clarify about relationship among the parameters how they are closed to each other to evaluate the actual prediction of health score. In our data analysis, we have shown Support vector machine i.e., SVM have the highest accuracy rate (near about to 93%) among the other chosen classifiers as well as simple neural networking approach. The result follows on accuracy, recall, precision and confusion matrix are also shown in this evaluation. Violin plot and two-dimensional kernel density estimation (2D KDE) plots are also implemented for visualization of the dataset.

1. Introduction

The healthcare sector is only now beginning to take off while cognitive computing is already making ripples in other areas. More individuals are benefiting from it, but it's also making it more difficult to manage the healthcare system since an increasing number of elderly people aren't making informed decisions about their health, which leads to a decline in care quality and an increase in healthcare problems [1]. One of the most pressing problems in healthcare is making sure all patients get the care they need without long wait periods or mountains of paperwork. The potential for machine learning as a solution to these issues has been identified. This study's recommended approach seeks to improve healthcare efficiency by developing an intuitive platform that streamlines communication between physicians and their patients and by utilizing machine learning algorithms to expedite the diagnostic procedure [2]. Due to a shortage of facilities, funds, and qualified medical personnel, developing nations may have restricted access to healthcare. One possible answer to this issue is the Portable Health Clinic (PHC), which may be readily relocated to places that lack adequate healthcare services. [3]. In spite of the widespread use of EHRs and EMRs, paper documents are still widely used in the healthcare industry. [4]. Figure 1 shows the Block diagram of EMR system. Some healthcare institutions, especially those in smaller clinics or rural locations, still use paper-based systems, even though many have adopted electronic medical records and electronic health records [5]. Electronic medical records (EMRs) are digital representations of patients' medical records kept by a single healthcare provider or institution. In addition to facilitating the acquisition and transmission of patient data within the hospital, this system also aids in the regulation of information flow during therapy. [36]. Despite the fact that EMR is a digital record, there are numerous issues that arise when it is used in healthcare settings, and very few of these are addressed [37]:

- Electronic medical records cannot be shared with anybody outside of the practice.
- When new information becomes available, patient medical histories from different healthcare

facilities won't be updated or made public.

- Why Since EMRs are kept centrally, patients frequently have to spend more time and money on several tests at different healthcare facilities.

Paper copies of electronic medical record (EMR) papers are still commonly used for internal company sharing, harkening back to a more conventional era.

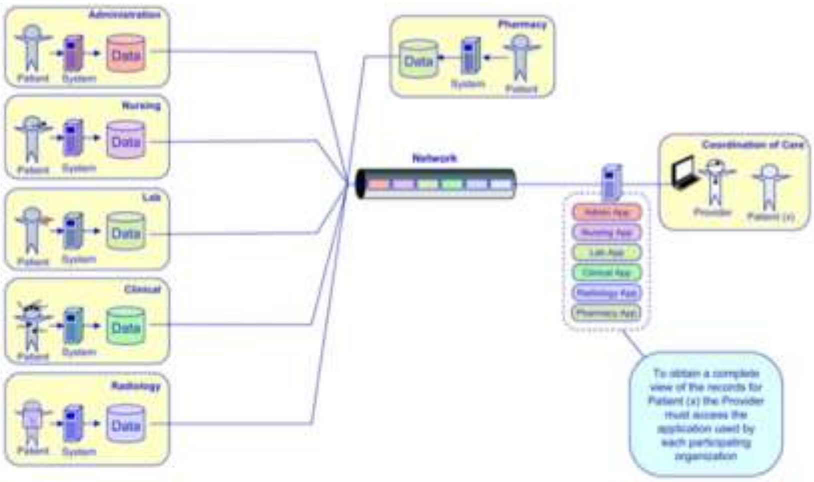


Figure 1: Block diagram of EMR system.

In response to these issues, several types of health information systems were developed, including EMR, PHR, and EHR [4]. EHRs, which are digital copies of patients' medical records, let healthcare institutions provide better patient care by providing features that other kinds of apps are unable to provide. EHR integration at all levels of the health system helps combat the issue of patient data mobility during the point of service since it facilitates and secures data flow [38]. An EHR system is defined as one that is "built to go beyond standard clinical data collected in a provider's office and can be inclusive of a border view of a patient's care" by the Office of Health Information Technology [39].

Making decisions is aided by any decision support system. When it comes to patient care, the decision-making process will function as a mentorship program for aspiring physicians and a secondary opinion system for specialists. The DSS, which is connected to the EHR system, has a wealth of patient data. Juniors, trainees, doctors, and researchers are helped by the DSS with EHR system to provide improved medical treatment, diagnosis, and determining the suitable level of health. Practitioners have access to EHR data from patients, labs, photo studios, and other sources anywhere, at any time, and on any device. Any decision support system must go through the entire process of finding and evaluating data via the EHR system as opposed to using the conventional approach. Computer-based applications are used in data analysis inside EHRs to offer guidelines

ISSN (ELECTRONIC): 2945-3135

Pramit Kumar Samant, Vinay Pathak & Waqar Ahmad (2026). Machine Learning Based Decision Support System on Electronic Health Record Data to Predict Health Score. International Journal of Multidisciplinary Research & Reviews, 5(7),239-274.

and prompts to help with prompt medical care. However, when medical professionals have access to precise and comprehensive data, this is feasible. With decision support systems, EHRs can enhance the capacity to anticipate, diagnose, and lessen or avoid medical mistakes, improving patient outcomes. Faster patient diagnosis is made possible by the decision support system's entire picture. In addition to keeping track of the patient's prescriptions and allergies, a trained EHR can automatically identify problems that may arise when a new medicine is prescribed and alert the doctor to any potential conflicts that might compromise the patient's care.

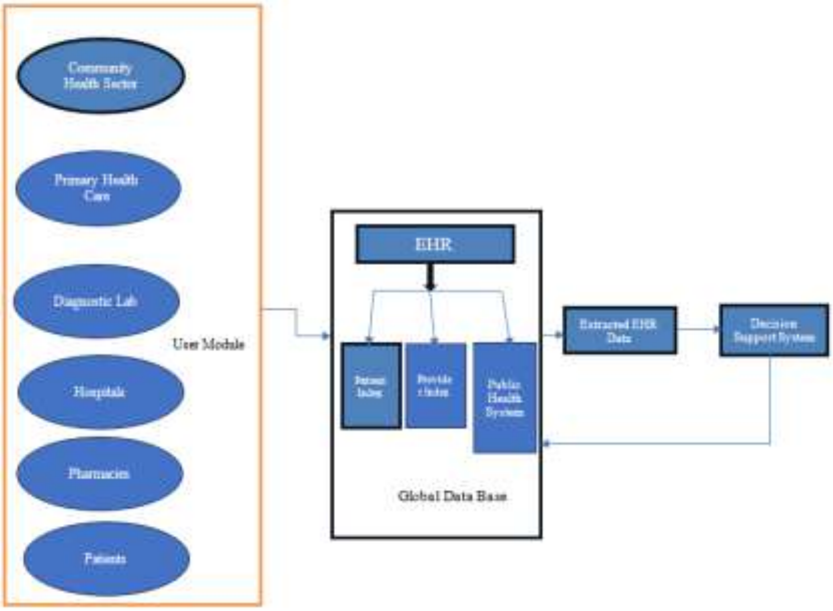


Figure 1: Block diagram of EHR System for health score prediction [32]

As shown in Figure 2 schematic of the specified EHR platform [32]. Modules for users, storage, data retrieval, and decision support system (DSS) make up the system in question. Information from many different areas of healthcare, including primary care, diagnostic laboratories, hospital pharmacists, patients, and community health workers, will make up the user module. The EHR storage module is responsible for storing the data in the database. The data extraction module then pre-processes and extracts the relevant data from the EHR. The researched approach's architecture is shown by the pre-processed data set that is provided to the decision support system.

1.1 Motivation

The impetus for creating a machine learning-based decision support system utilizing EHR data arises from the urgent necessity to enhance healthcare outcomes and tackle significant issues in healthcare delivery. The contemporary healthcare environment confronts various challenges, including the rising prevalence of chronic diseases, soaring healthcare expenses, and the necessity for high-quality, prompt care. Machine learning-based solutions facilitate data-driven decision-making, thereby optimizing workflows, delivering personalized healthcare insights, and minimizing human error. Implementing such a system enables healthcare organizations to adopt a proactive approach to patient health management, as opposed to the conventional reactive paradigm. This transition not only aims to enhance patient outcomes but also reduces the burden on healthcare providers by automating standard analyses and issuing alarms for potential dangers.

Furthermore, the ability of machine learning algorithms to manage extensive datasets and discern intricate patterns is crucial for forecasting health scores and evaluating patient risk. EHRs encompass extensive information, including patient demographics, medical history, laboratory findings, and vital signs, which combined provide a holistic perspective of a patient's health. Machine learning algorithms, including Support Vector Machines (SVM) and artificial neural networks (ANN), can proficiently analyze this data to uncover latent trends and correlations that may elude human therapists. These insights provide more precise health risk evaluations, directing treatments prior to the deterioration of a patient's condition. The predictive capability is particularly beneficial in chronic illness management, where early identification and intervention are crucial for treating disorders such as diabetes, heart disease, and hypertension.

This technology fulfils a vital requirement in healthcare for data integration and interoperability. Historically, healthcare data is disjointed and dispersed across multiple systems, complicating clinicians' ability to obtain a comprehensive understanding of a patient's health. Employing machine learning techniques to standardize and integrate this data, a decision support system grounded in electronic health records may address these deficiencies, guaranteeing that each data point significantly enhances patient care. This not only improves clinical decision-making but also facilitates continuity of care, allowing patient information to transition smoothly among various healthcare practitioners and environments. The creation of a machine learning-driven decision support system utilizing EHR data corresponds with the global movement towards preventive and personalized healthcare. As populations expand and age, global healthcare systems are experiencing heightened pressure. In response, there is a significant emphasis on preventive care, which seeks to proactively manage patient health and avert illnesses prior to their onset. Machine learning-based solutions enhance this preventative strategy by forecasting health scores and offering early alerts, thereby assisting individuals in preserving optimal health and circumventing expensive procedures in the future. Moreover, the accuracy of these forecasts facilitates personalized care, customized to the distinct health profile of each patient, so improving the overall quality of healthcare.

1.2 Novel Scientific Contribution

The investigation on a Machine Learning-Based Decision Support System for EHR data makes significant scientific contributions, especially in enhancing healthcare analytics and predictive modelling. This study presents a novel implementation of Support Vector Machine (SVM) technology within an Electronic Health Record (EHR) system, which adeptly assists healthcare practitioners in making data-informed decisions based on patient health scores. The study excels in analysing extensive EHR datasets (exceeding 79,000 records) by combining machine learning with exploratory data analysis (EDA) approaches, hence improving the predicted accuracy of health status evaluations. Thorough comparative research reveals that the SVM model attains a higher accuracy (~93%) in predicting health ratings, surpassing other classifiers and neural network methodologies. The study underscores the significance of data visualization techniques, including violin plots and two-dimensional kernel density estimation (2D KDE), which enhance comprehension of data distributions and correlations across health markers. We summarize the key contribution as follows

- Development of an SVM-based decision support system for health score prediction using EHR data, achieving high predictive accuracy.
- Comparative evaluation demonstrating SVM's superior performance over other classifiers and neural networks in healthcare analytics.
- Integration of EDA and data visualization techniques to better interpret EHR data and improve clinical decision-making.

2. Literature Survey

A literature review on EHR-based decision systems is essential as it offers an in-depth comprehension of current approaches, difficulties, and progress in healthcare analytics. Examining previous research enables scholars to discern deficiencies in existing models, such constraints in predicting precision, data amalgamation, or interpretability, and to enhance effective methodologies. It emphasizes the distinct requirements of healthcare systems concerning data privacy, interoperability, and clinical adoption, hence facilitating the creation of more efficient, secure, and broadly embraced systems. This fundamental information ultimately facilitates the development of new, effective solutions that enhance patient care and health outcomes.

2.1 EHR based Decision System

E-health solutions that use machine learning have a lot of potential. Large datasets can be analyzed by machine learning algorithms, which can then spot trends and anticipate outcomes for better, more individualized care. As per [6], for activities like disease detection, therapy suggestions, and patient monitoring, these systems have been frequently used. The ability of machine learning-based decision support systems to manage enormous amounts of patient data, which may be used for early disease identification and individualized treatment regimens, is one of their main advantages [7]. These systems’ ability to interpret both structured and unstructured data, including text and natural language, medical pictures, and electronic health records (EHRs), will increase the precision of diagnoses.

Despite the potential advantages, there are a number of difficulties in implementing machine learning-based decision support systems in e-health. First, there are issues with data security and privacy. Sensitive information is contained in healthcare data, so protecting that information is crucial [8]. The Health Insurance Portability and Accountability Act (HIPAA) in the US is one example of a stringent rule that must be complied with, adding complexity to system construction.

The interoperability and quality of the data are another important difficulty. Healthcare data is frequently dispersed across numerous systems and kept in various formats [9]. The incorporation of machine learning algorithms may be hampered by this lack of standardization because these algorithms depend on accurate and reliable data to function [10]. In order to allow effective data management, healthcare institutions must also invest in infrastructure and data storage capabilities. Clinical acceptability is another difficulty. Machine learning-based recommendations may not be widely adopted because healthcare practitioners may be reluctant to trust them [11]. It is crucial to involve physicians in the development process and make sure that these technologies enhance rather than replace their knowledge.

Telemedicine, e-health and digital dermatology are all terms used to describe the examination and treatment of patients using apps and software platforms. The gathering of patient information, the transmission of medical records, the identification of the ailment, and the creation of treatment plans are all made easier by telemedicine and e-health technology [12]. Future decisions are therefore based on past choices and the results of those choices. E - Healthcare Administration Framework is planned for multispecialty healing centers, to cover a wide run of healing center organization and administration forms [13]. It is a coordinate's end-to-end E - Healthcare Administration Framework that gives pertinent data over the clinic to back successful decision making for quiet care, healing center organization and basic monetary bookkeeping, in a consistent stream [14]. Potential advantages and uses.

There are many uses for machine learning-based decision support systems in e-health. They are particularly good at diagnosing conditions from medical imaging, forecasting patients' prognoses, and prescribing treatments. By offering in-the-moment alerts and information, these devices can also assist clinical decision-making [15]. Additionally, by maximizing resource allocation, reducing administrative duties, and avoiding medical errors, they have the potential to lower healthcare expenditures. Predictive analytics, for instance, can assist hospitals in effectively allocating staff, lowering patient wait times. Telemedicine improvement is a substantial additional benefit [16]. Studying the efficacy of machine learning models for automated diagnosis [33] proved that machine learning models could effectively perform autonomous diagnosis. Results from an Android app were used to discuss Coronavirus, diabetes, and cardiovascular disease in the study. On a real-time website, a pre-trained model was used to assess and process the data. The application of machine learning in the diagnosis of human diseases is comparable to this study[34] regarding the medical industry's data tools. A piece of literature [35] found that with 10 neighbors, KNN achieved an accuracy of around 86% on the training dataset and roughly 77% on the testing data. An electronic health record (EHR) that has been educated may automatically identify problems with a new prescription and alert the doctor to any potential conflicts, all of which contribute to the patient's health monitoring [41]. It is also capable of monitoring the patient's allergy and drug history. Furthermore, by offering clinical alerts and reminders, improving patient data integration, analysis, and communication, facilitating consideration of all aspects of a patient's condition, integrating safeguards against adverse event prescription, and enhancing research and monitoring to enhance clinical quality, the EHR also helps better disaster management [42]. A knowledge base system works by constructing rules, which are essentially if-then statements. The system then uses the returned data to test the law and either generates action or output. [43]. Proof based on patient records, performance, or both can inform regulations [44]. While data sources are still necessary for non-knowledgeable DSS, the focus shifts from following professional medical information to improving AI, ML, or mathematical pattern recognition.[43]. Concerns with data availability and the difficulty in comprehending the notion of how it employs AI to generate suggestions are among the numerous issues brought about by the fast expansion of AI in non-knowledgeable DSS applications [44]. Among the many services offered by DSS are diagnostics, alarm systems, medication control, prescription, disease management, and a whole lot more. [45].

Table 1. Comparative Analysis of existing models.

Reference	Key Objective/Feature	Methodology	Strength	Limitation	Complexity
[6]	for activities like disease detection, therapy suggestions, and patient monitoring, these systems have been frequently used.	The ability of random forest decision support systems to manage enormous amounts of patient data.	Effectiveness in to manage enormous amounts of patient data.	Delay in response time, less scalable, 78% accuracy	High complexity in implementation and management
[7]	These systems' ability to interpret both structured and unstructured data, including text and natural language, medical pictures, and electronic health records (EHRs), will increase the precision of diagnoses.	The ability of linear regression decision support systems to manage enormous amounts of patient data.	The assumption of linearity between variables restricts, reduced False Positive Rate (FPR), False Negative Rate (FNR), and Average Energy Consumption (AEC) compared to existing algorithms	high energy consumption, less scalable, potential delays in response time and sensitive to outliers, 63% accuracy	Complexity associated with structured and unstructured data

Pramit Kumar Samant, Vinay Pathak & Waqar Ahmad (2026). Machine Learning Based Decision Support System on Electronic Health Record Data to Predict Health Score. International Journal of Multidisciplinary Research & Reviews, 5(7),239-274.

[8]	the potential advantages, there are a number of difficulties in implementing Bayesian-based decision support systems in e-health.	The Health Insurance Portability and Accountability Act (HIPAA) in the US is one example of a stringent rule that must be complied with, adding complexity to Bayesian based decision system construction.	Improved detection accuracy, to manage enormous structured and unstructured data	no general way to represent and handle the uncertainty within the background knowledge and the prior probability function, 71% accuracy	Moderate computational overhead
[9]	The interoperability and quality of the data are another important difficulty. Healthcare data is frequently dispersed across numerous systems and kept in various formats	The ability of logistic regression-based decision support systems to manage enormous amounts of patient data with managing various format	Improved detection accuracy, to manage various format	Limited scalability due to number of observations is lesser than the number of features, If you use a linear format for logistic regression, it may affect the data, 74% accuracy	Moderate due to constructs linear boundaries
[10]	The incorporation of machine learning may be hampered by this lack of standardization because these algorithms depend on accurate and reliable data to function	The ability of Nave-Bayes based decision support system to manage standardization	Improves reliable data to detect diseases with high throughput	Lack of scalability in large-scale of datasets, 76% accuracy	Moderate due to quality of data
[15]	Machine learning is particularly good at diagnosing conditions from medical imaging, forecasting patients' prognoses, and prescribing treatments. By offering in-the-moment alerts and information, these devices can also assist clinical decision-making	Lasso regression-based decision support system to forecast patient prognoses, to make security enhancement	Improved detection accuracy with high throughput	May face outliers in large-scale data,67 % accuracy	Moderate due to cross validation technique to find the value can be as expensive as stepwise selection techniques
[16]	Predictive analytics, for instance, can assist hospitals in effectively allocating staff, lowering patient wait	The ability of decision tree support systems to manage enormous amounts of patient	Effectiveness in to manage enormous amounts of patient data.	Delay in response time, less scalable, 78% accuracy	Moderate due to small change in the data can result in a major change in the structure of

Pramit Kumar Samant, Vinay Pathak & Waqar Ahmad (2026). Machine Learning Based Decision Support System on Electronic Health Record Data to Predict Health Score. International Journal of Multidisciplinary Research & Reviews, 5(7),239-274.

	times. Telemedicine improvement is a substantial additional benefit.	data and additional benefit.			the decision tree, which can convey a different result from what users will get in a normal event.
[33]	Machine learning based decision support system was used in the diagnosis of human diseases on the data tools used in the medical field.	The ability of random forest decision support systems to manage enormous amounts of patient data.	Improved detection accuracy with label data on existing algorithm	high energy consumption, less scalable, potential delays in response time and decision-making processes, 89% accuracy	Moderate due to a large number of trees can make the algorithm too slow and ineffective for real-time predictions
[34]	machine learning was used in the diagnosis of human diseases	The ability of multi linear regression decision support systems to manage enormous amounts of patient data.	Improved in feature selection and training and test fit selection	high energy consumption, potential delays on training data and decision-making processes, 81% accuracy	Moderate due to small multiple correlation coefficient, unstable regression weight
[35]	One of the works showed that KNN for 10 neighbors provided an accuracy of about 86% for the training dataset and about 77% for testing data.	Utilizes intelligent K Nearest Neighbors classification, feature selection	Improved accuracy rate i.e. 86% on the test dataset.	Communication and computational overheads, real-time implementation challenges, scalability issues, accuracy 87% in test data	Moderate due to scalability and training process
[41]	An electronic health record (EHR) that has been educated may automatically identify problems with a new prescription and alert the doctor to any potential conflicts, all of which contribute to the patient's health monitoring	The ability of naive bays decision support systems to manage enormous amounts of patient data	Improved accuracy on the test datasets and cross validation for beta testing	Communication and computational overheads, outliers and noise in data81% accuracy rate	Moderate due to quality of data like outlier and noise
[42]	It is also capable of monitoring the patient's allergy and drug history. EHR also helps better disaster management	The ability of decision tree-based decision support systems to manage enormous amounts of patient data.	Improved accuracy on the test datasets and high throughput	Communication and computational overheads, real-time implementation challenges, 83% accuracy rate	Moderate due to scalability and training process
[44]	Regulations can be made using	The ability of decision tree-	Effectiveness on accuracy to	Remote station and computational	High complexity

	documentary, performance-based, or patient-based evidence	based decision support systems to manage enormous amounts of patient data.	manage enormous amounts of patient data.	overheads, real-time implementation challenges, 81% accuracy rate	
DSS (Proposed)	To make novel Decision support system for EHR, To explore EHR data manipulation via machine learning algorithms, focusing on observer benefits, To compare SVM and ANN for EHR dataset for accuracy purpose, also, To identify key dataset attributes for healthcare score prediction.	The ability of support vector machine-based decision support systems to detect health status. Also, comparing health scores between ANN and SVM. Finding the most important attributes with EDA technique	To detect health score novel model, comparing the accuracy between two model, finding best attributes to find health score, accuracy rate of SVM 93%	SVM based DSS is suitable for large datasets with many features	Moderate due to quality of data and high complexity

The survey found that decision assistance systems based on support vector machines are suitable for every person's health status. Many healthcare information systems can also benefit from prediction systems' decision-making capabilities. The research inspired us to create an innovative SVM-based decision support system for the EHR and train it to detect health scores from the 79,540 records of health data extracted from the EHR. To determine which machine learning technique yields the greatest results, we will assess how well the proposed architecture works. To find any possible limitations or areas for development, the provided architecture will be tested using real-world datasets and various scenarios.

3 Proposed Method

The proposed work is divided into various subsections as follows.

3.1 Support Vector Machine Based Decision Support System

The proposed SVM based DSS system is shown in the mentioned Figure 3. This module contains seven components. EHR system is the first component to provide 79540 records. The second component is data preprocessing to clear noise on data. The third component is input and output extraction to separate input features and output feature separately. Forth component is data scaling to scale the data between 0 and 1. After scaling, input data and output data are used for cross validation. Now training data sets are passed on to support vector machine to build model. Thereafter, trained model is tested by test data to detect health status with 93% accuracy.

ISSN (ELECTRONIC): 2945-3135

3.2 Support Vector Machine Algorithm

For regression and classification purposes, supervised learning models with associated learning methods can be found in Support Vector Machines (SVMs). Cortes & Vapnik's SVM approach [16] finds the optimal hyperplane in the space of n-dimensional classifications that maximizes the distance between classes. It is commonly believed that SVM outperforms other classifiers [17], Noted, however, is the fact that the main rationale for opting for an SVM rather is the possibility that the problem is not linearly separable. [18]. A support vector machine (SVM) trained with a non-linear kernel, like the Radial Basis Function (RBF), could be the best option in this case. If one is in a space with high dimensions, SVMs are also useful. While this requires a lot of training time, SVMs have proven to work better for text categorization. Figure 2 shows the SVM algorithm in 2-dimensions. An expansion of the support vector classifier, the support vector machine (SVM) is born out of a specific way the feature space grows with the help of kernels [19]. Eq. (1) gives the representation of a linear support vector classifier (1):

$$f(x) = \beta_0 + \sum_{i=1}^{n} \alpha_{i\langle x, x_i \rangle} \qquad (1)$$

where the parameters α1, ⋯, αn, and β0 are estimated using $\binom{n}{2}$ inter products ⟨xi,x′i⟩ among all sets of training data. replacing the inner product with K(xi,x′i), where K is a kernel function. Eq. (2) shows how the linear kernel is represented.

$$K(x_i, x_i^{'}) = \sum_{j=1}^{p} x_{ij} x_{ij}^{'} \qquad (2)$$

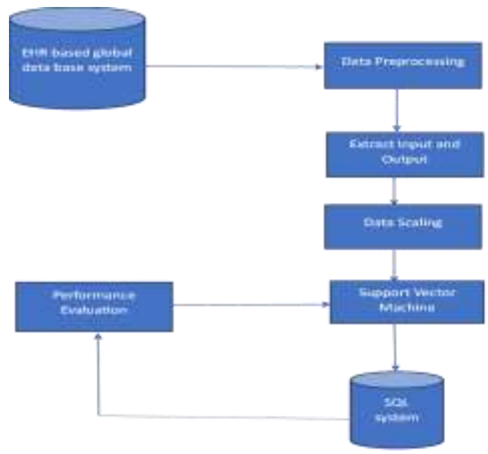


Figure 3: Proposed SVM based Decision Support System for EHR based System.

Pramit Kumar Samant, Vinay Pathak & Waqar Ahmad (2026). Machine Learning Based Decision Support System on Electronic Health Record Data to Predict Health Score. International Journal of Multidisciplinary Research & Reviews, 5(7),239-274.

Polynomial kernel of degree d (where d is positive) may be expressed as illustrated in Eq. (3):

$$K\left(x_i, x_i^{'}\right) = \left(1 + \sum_{j=1}^{p} x_{ij} x_{ij}^{'}\right)^d \qquad (3)$$

Classification outcomes from a combination of non-linear kernel and support vector classifier are termed the SVM. Using a large number of training instances is not recommended for the SVM classifier because it is a kernel-based supervised learning method that was designed for binary classification and divides data into two or more groups. To make the training set more like a linearly separable data set, a mapping technique called a kernel function is applied. Effective use of a kernel function allows mapping to achieve its purpose of increasing the data collection's dimensionality. Linear, RBF, quadratic, multilayer perceptron, and polynomial kernels are among the most popular types of kernels [20]. While RBF kernel function excels with non-linear data sets, linear kernel function shines with linearly separable data sets. In comparison to the RBF kernel function, the linear kernel function reduces the amount of time needed to train the SVM. Additionally, overfitting is less common with the linear kernel function than the RBF kernel function. The regularization parameter C, sometimes called the box constraint, and the kernel parameter, often called the scaling factor, determine the SVM classifier's performance. When used together, they form what is called the hyperplane parameter [21].

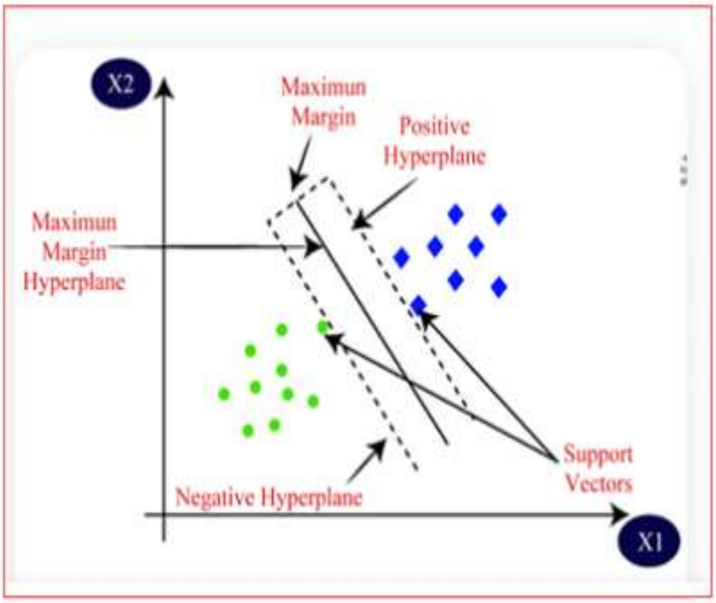


Figure 4: SVM algorithm in 2-dimensions.

Pramit Kumar Samant, Vinay Pathak & Waqar Ahmad (2026). Machine Learning Based Decision Support System on Electronic Health Record Data to Predict Health Score. International Journal of Multidisciplinary Research & Reviews, 5(7),239-274.

During training, support vector machines (SVMs) create a model, draw a boundary map for each class, and determine the hyperplane that separates the classes. To improve the accuracy of classification, it is helpful to increase the hyperplane margin, which increases the distance between the classes. Non-linear classification can also be done efficiently with SVMs [22]. Among the several varied domains where SVMs have found fruitful use are text and hypertext classification, picture identification, validation, and recognition [23], fields such as bioinformatics, protein classification, computational chemistry, speech recognition, bankruptcy prediction, time series forecasting, information retrieval, and remote sensing image analysis data from spectroscopy, for instance, NMR and chromatography-mass spectrometry

Algorithm SVM Workflow

1. **Input:** Raw Dataset (EHR Data)
2. **Output:** Trained SVM Model, Model Evaluation Metrics
3. **Step 1: Data Normalization**
4. Normalize the dataset
5. *NoramalizedData* ← *Normalize (RawDataSet)*
6. **Step 2: Kernal Transformation**
7. Choose Kernal Function (e.g., Linear, polynomial, RBF)
8. Transform the normalized dataset using the Kernal function
9. *KernalData* ← *ApplyKernal (NormalizedData, KernalFunction)*
10. **Step 3: Hyperplane Optimization**
11. Define optimization problem to find the optimal hyperplane
12. Solve the optimization problem
13. *OptimalHyperplane* ← *OptimizeHyperplane(KernalData)*
14. **Step 4: Cross-Validation Training**
15. Split *KernalData* into training and validation sets
16. **for** each fold in Cross-Validation **do**
17. Train the SVM model on the training set
18. Validate the model on the validation set
19. Adjust hyperparameters based on validation results
20. **end for**
21. *TrainedModel* ← *TrainSVM Model(KernalData)*
22. **Step 5: Prediction Simulation**
23. Apply the same kernal function to new test data
24. *KernalTestData* ← *ApplyKernal(TestData, KernalFunction)*
25. Use *TrainedModel* to make predictions
26. *Prediction* ← *Predict*(*TrainedModel, KernalTestData*)
27. **Step 6: Statistical Indicators**

ISSN (ELECTRONIC): 2945-3135

Pramit Kumar Samant, Vinay Pathak & Waqar Ahmad (2026). Machine Learning Based Decision Support System on Electronic Health Record Data to Predict Health Score. International Journal of Multidisciplinary Research & Reviews, 5(7),239-274.

28. Compare *Predictions* with true labels of test data
29. Calculate performance metrics (Accuracy,Precision, Recall, F1 Score, etc.)
30. *Metrics* ← *EvaluateModel(Predictions, TrueLabels)*
31. Display or log *Metrics*
32. **Step 7: End**
33. Conclude the process
34. **Output:** *TrainedModel, Metrics*

3.2.1 Kernal mathematical intuition

In the non-linear data set, dual problem is mentioned in the form of equation for n features

Dual problem $D(g) = \sum_{i=1}^{m} g_i - \sum_{i=1}^{m}\sum_{k=1}^{m} g_i g_k \, [y^i y^k * x^i x^k]$

Where $x^i x^k$ is dot product of features

The non-linear function $f\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1^2 \\ \sqrt{2x_1x_2} \\ x_2^2 \end{pmatrix}$

$$f(a) = \begin{pmatrix} a_1^2 \\ \sqrt{2a_1}a_2 \\ a_2^2 \end{pmatrix}$$

$$f(b) = \begin{pmatrix} b_1^2 \\ \sqrt{2b_1}b_2 \\ b_2^2 \end{pmatrix}$$

$$f(a)\,f(b) = \begin{pmatrix} a_1^2 \\ \sqrt{2a_1}a_2 \\ a_2^2 \end{pmatrix} . \begin{pmatrix} b_1^2 \\ \sqrt{2b_1}b_2 \\ b_2^2 \end{pmatrix}$$

$$= a_1^2 b_1^2 + 2a_1b_1a_2b_2 + a_2^2 b_2^2$$

$$= (a_1b_1 + a_2b_2)(a_1b_1 + a_2b_2)$$

$$= (\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}\begin{pmatrix} b_1 \\ b_2 \end{pmatrix})(\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}\begin{pmatrix} b_1 \\ b_2 \end{pmatrix})$$

$$f(a)\,f(b) = (a.b)(a.b)$$

Therefore, The Kernal function $K(a, b) = (a.b)^2$

Here RBF Kernal function is used to classify the EHR-based non-linear data set more accurately i.e

$$K(x_1 x_2) = e(|x_1 - x_2|^2 / 2a^2)$$

Where a is standard deviation for two features which has been mentioned.

Therefore, the significance of kernal function to find optimal separating hyperplane on EHR dataset to enhance SVM algorithm in high dimension space.

3.2.2 multi-class classification (normal to sever) mathematical intuition

Initially, the multi-classification problem is breaking down into smaller subproblem. Now, subproblems appear as binary classification problems. Then, apply binary classification principles for classifying all smaller subproblems. The majority voting or highest value of the sub problems output will predict result. Here, one vs Rest (OVR) approach is used i.e.

clf=svm.SVM(decision_function_shape = ”ovr”)

In OVR, assume N class problem, then need N SVM classifiers as follows:

SVM classifier_1 learns “class_output=1” vs “class_output=other”

SVM classifier_2 learns “class_output=2” vs “class_output=other”

SVM classifier_3 learns “class_output=2” vs “class_output=other”

.

.

.

SVM classifier_N learns “class_output=N” vs “class_output=other”

However, OVR approach uses n-SVM classifier only.

Therefore, the significance of multi-class classification approach i.e. OVR approach can learn nonlinear decision boundaries to make them suitable for EHR dataset.

3.2.3 The flow chart of SVM algorithm and workflow of the algorithm

The flow chart of SVM algorithm to be shown in figure 5. Now, three domain factors—administrator, doctor, and patient—make up the user module. The value of all the mentioned attributes like temperature, respiration system, systolic blood pressure, diastolic blood pressure and pulse oximeter to be measured by the SVM algorithm.

Pramit Kumar Samant, Vinay Pathak & Waqar Ahmad (2026). Machine Learning Based Decision Support System on Electronic Health Record Data to Predict Health Score. International Journal of Multidisciplinary Research & Reviews, 5(7),239-274.

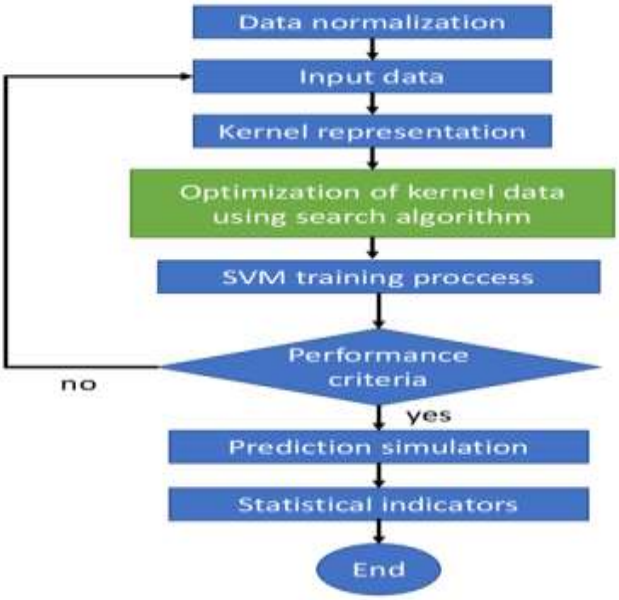


Figure 5: Flow chart of support vector machine algorithm.

Support vector machine algorithm follows normalization, kernel representation, optimization of kernel, performance criteria to be followed, prediction simulation and statistical indicators to be mentioned in the flow chart to be shown in figure 5.

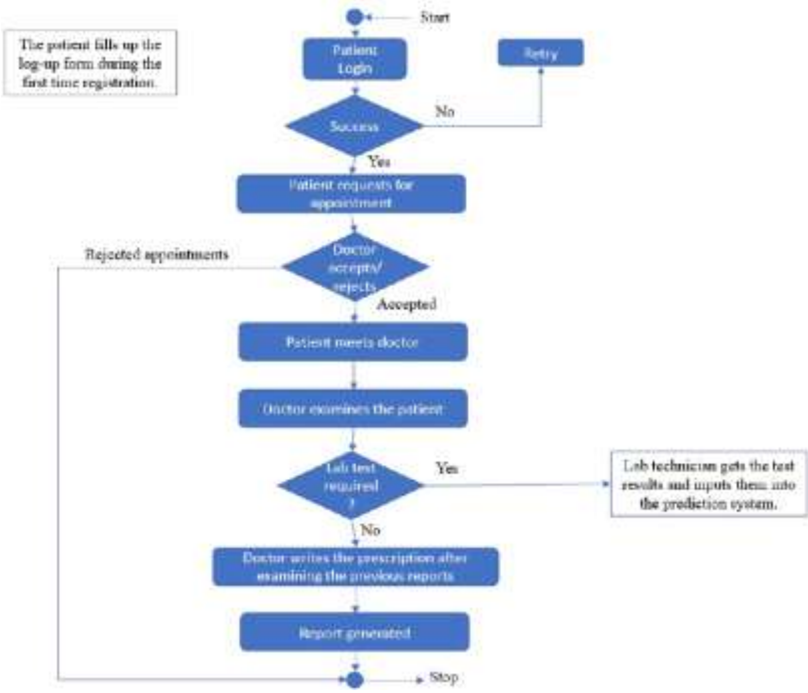


Figure 6: Working model of the EHR system for health score prediction.

Pramit Kumar Samant, Vinay Pathak & Waqar Ahmad (2026). Machine Learning Based Decision Support System on Electronic Health Record Data to Predict Health Score. International Journal of Multidisciplinary Research & Reviews, 5(7),239-274.

As flow chart to be shown in figure 6 measures health scores with the help of attributes like temperature, respiration, systolic blood pressure, diastolic blood pressure and pulse oximeter taken from stored lab test. The suggested system's workflow integrates all three domains is shown in the figure for health score prediction.

4. Theoretical Analysis and Results Discussion

Machine learning algorithms are computer programs that use math to discover hidden patterns in data. Machine learning is a set of computer algorithms that can identify patterns, categorize things, and make predictions using data that it has learned from before [24]. The most used methods in biology to learn and make predictions are support vector machines (SVM) and artificial neural networks (ANN). We can use a type of computer model called ANN to predict certain characteristics of DNA and RNA binding proteins, non-coding variants, alternative splicing, and the effectiveness of drugs [25]. Deep learning models are better than other machine learning methods at finding complicated things in data [26].

Here are some machine learning algorithms that are shortly discussed [27].

Linear Regression - Linear regression is a basic method used to show the straight-line connection between input features and a continuous target variable [28]. It works by drawing a straight line through the information and using that line to guess what other values might be.

Logistic Regression - Logistic regression is a type of analysis that helps us determine the probability that a given thing belongs to a certain category [29]. It is used for classifying things.

SVM (Support Vector Machine) - SVMs are algorithms that help us classify and predict data in a supervised learning way [30]. It looks for a straight line that separates different groups of things in a space with different features.

KNN (K-nearest Neighbor) - KNN is a method that can be used to classify or predict things without requiring certain conditions or assumptions [31]. To put it simply, it finds data points that are very similar to a new data point and uses those points to make a label prediction for the new data point.

Decision Tree - Decision trees are a type of learning method that can be used to classify things and make predictions about values. It works by breaking down the data into smaller groups until each group can be accurately classified or predicted.

Random Forest - Random forests are a method of learning that use multiple decision trees to make predictions by combining the predictions of each tree together. It makes single decision trees more accurate and stronger. It can be used for sorting things into groups or for predicting values.

Naive Bayes - Naive Bayes is a type of classifier that uses probabilities to make predictions in classification tasks. This method assumes that the characteristics of a piece of data do not rely on one another.

The data collections of this study are crucial for demonstration of the evaluation of the project's objectives. The data collection should be in a manner where a researcher can be able to evaluate the project's outcomes through detailed analysis. In the dataset, medical parameters are crucial to observe for evaluating the health status as in health score.

4.1 EHR Data Set Collection

The medical EHR dataset was collected from the Kaggle where mentioned attributes like temperature, respiration system, systolic blood pressure, diastolic blood pressure and pulse oximeter to be introduced. Showing the dataset's nature and its characteristics:

```
dataset.head(79540)
```

	TEMPF	PULSE	RESPR	BPSYS	BPDIAS	POPCT	SCORE
0	99.1	90	16	129	75	99	0
1	97.5	71	16	167	82	98	0
2	98.4	89	20	118	76	98	0
3	97.6	85	16	124	95	98	0
4	98.2	80	18	156	92	98	0
...	...	...	...	...	...	...	...
79535	98.2	96	15	147	81	98	1
79536	98.7	83	21	161	87	96	2
79537	97.5	125	26	86	45	81	3
79538	98.0	128	16	127	96	99	2
79539	98.2	82	16	137	98	95	1

79540 rows × 7 columns

Figure 7: Characteristics and nature of EHR dataset

We have seen that there are seven columns and 79540 rows in the EHR dataset to be indicated the nature of the data in the figure 7. Meanwhile, these attributes to be observed during measuring health score status remotely.

TEMPF - Body temperature in Fahrenheit

PULSE -normally 60 bps to 100 bps

RESPR - normally range 0 to 12

BPSYS - normally range 120 to 129 mm Hg

Pramit Kumar Samant, Vinay Pathak & Waqar Ahmad (2026). Machine Learning Based Decision Support System on Electronic Health Record Data to Predict Health Score. International Journal of Multidisciplinary Research & Reviews, 5(7),239-274.

BPDIAS - normally range 80 to 89 mm Hg

POPCT - PULSE OXIMETER PCT oxygen saturation level (95% to 100%)

SCORE - Labeled as from 0 to 3 i.e. normal to severe

```
 #   Column   Non-Null Count  Dtype
---  ------   --------------  -----
 0   TEMPF    79540 non-null  float64
 1   PULSE    79540 non-null  int64
 2   RESPR    79540 non-null  int64
 3   BPSYS    79540 non-null  int64
 4   BPDIAS   79540 non-null  int64
 5   POPCT    79540 non-null  int64
 6   SCORE    79540 non-null  int64
dtypes: float64(1), int64(6)
memory usage: 4.2 MB
```

Figure 8: Mentioned Data Types of the EHR Attributes

We have also seen that all columns are non-null that means no empty value is in the dataset and all columns are int64 data type except TEMPF is in float64 data type to be shown in figure 8. It means no noise in the attributes i.e. to be cleaned in EHR dataset.

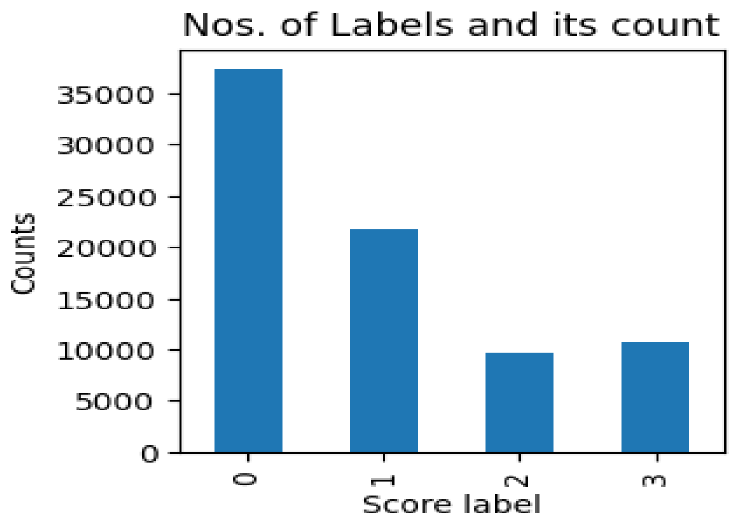


Figure 9: Categorization of Health Score

Pramit Kumar Samant, Vinay Pathak & Waqar Ahmad (2026). Machine Learning Based Decision Support System on Electronic Health Record Data to Predict Health Score. International Journal of Multidisciplinary Research & Reviews, 5(7),239-274.

Health score was categorized in between 0 to 3 in the range i.e. normal to severe to be shown in figure 9. Ratio among these health score categorization was 37412: 21700: 9704: 10724.it means 37412 to be considered as good health score i.e. 0, 21700 to be considered as low severe i.e. 1, 9704 to be considered as medium severe i.e. 2, and 10724 to be considered as worst risk i.e. 3.

```
dataset.describe()
```

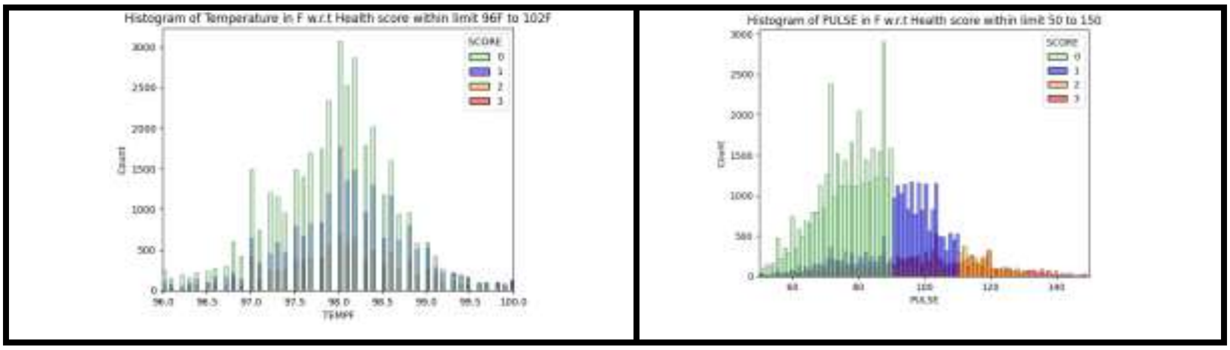
	TEMPF	PULSE	RESPR	BPSYS	BPDIAS	POPCT	SCORE
count	79540.000000	79540.000000	79540.000000	79540.000000	79540.000000	79540.000000	79540.000000
mean	98.072934	86.288257	18.646568	135.939955	79.046015	97.741463	0.921297
std	1.010266	17.373582	3.873378	23.452158	14.572607	2.809677	1.060885
min	89.500000	7.000000	8.000000	54.000000	21.000000	23.000000	0.000000
25%	97.600000	74.000000	16.000000	120.000000	70.000000	97.000000	0.000000
50%	98.000000	85.000000	18.000000	134.000000	79.000000	98.000000	1.000000
75%	98.500000	97.000000	20.000000	148.000000	88.000000	99.000000	2.000000
max	105.400000	200.000000	90.000000	280.000000	177.000000	100.000000	3.000000

Figure 10: Dataset's Statistics (similarity and dissimilarity)

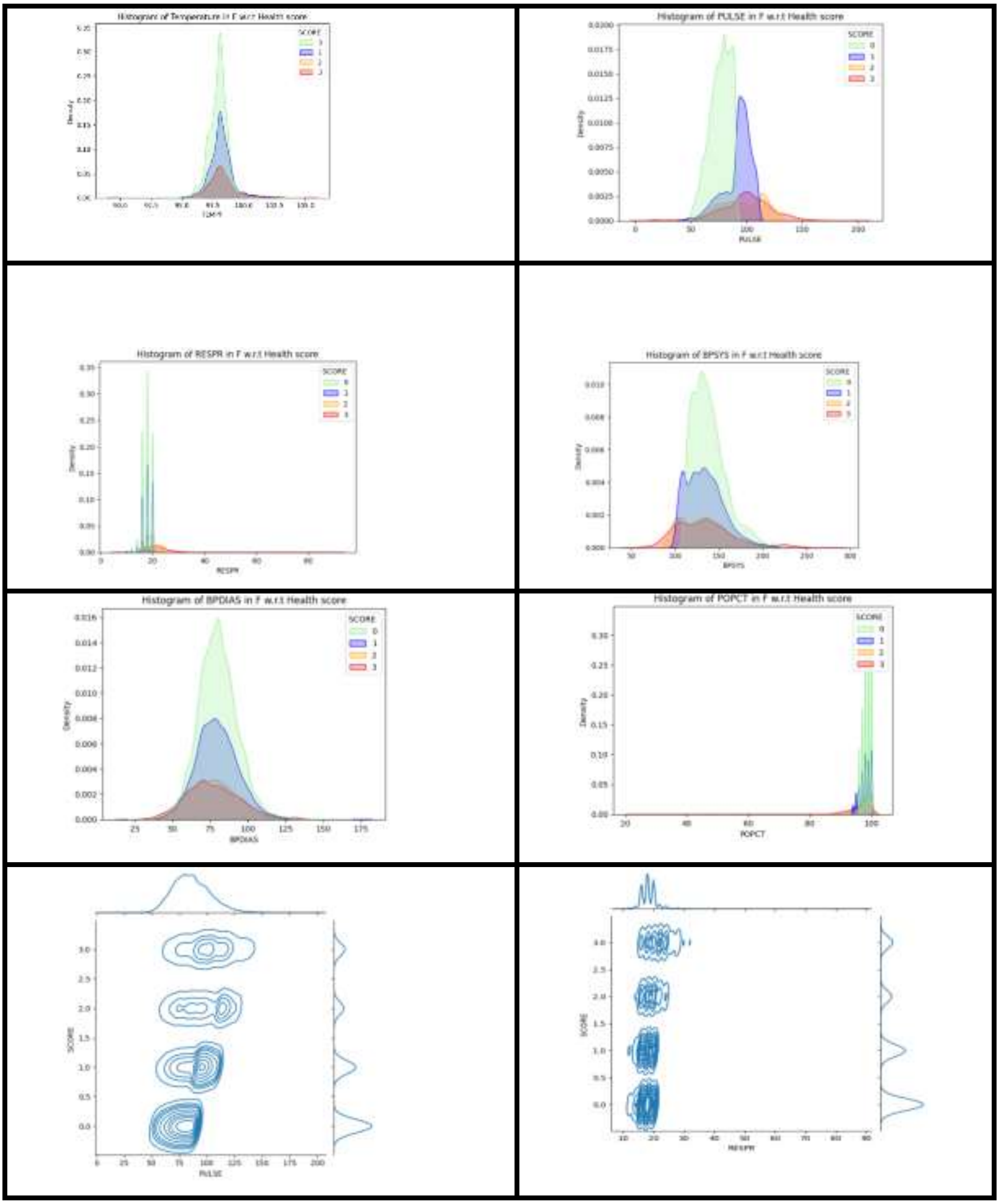
Therefore, we have also checked the dataset's statistics. Therefore, all columns have statistical behaviors to be shown in figure 10 with mention command dataset.describe() that will helpful to measure similar or dissimilar in EHR dataset.

4.2 Exploratory Data Analysis (EDA)

EHR data sets are normalized and scaled on Jupiter notebook with sklearn library that are passed on EDA for visualizing hidden patterns among TEMPF, PULSE, RESPR, BPSYS, BPDIAS, POPCT with SCORE to be shown in the below table 1.



Pramit Kumar Samant, Vinay Pathak & Waqar Ahmad (2026). Machine Learning Based Decision Support System on Electronic Health Record Data to Predict Health Score. International Journal of Multidisciplinary Research & Reviews, 5(7),239-274.



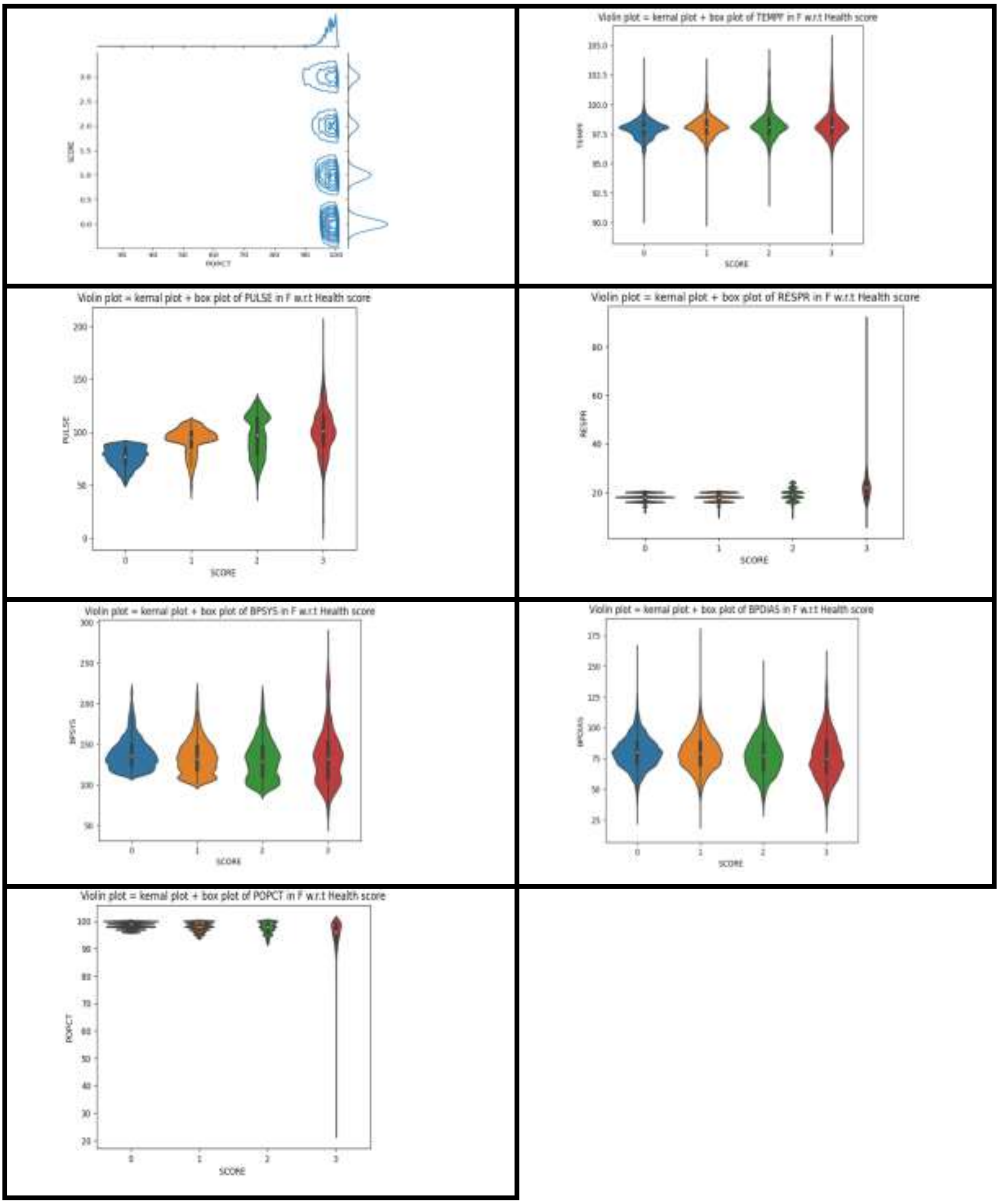


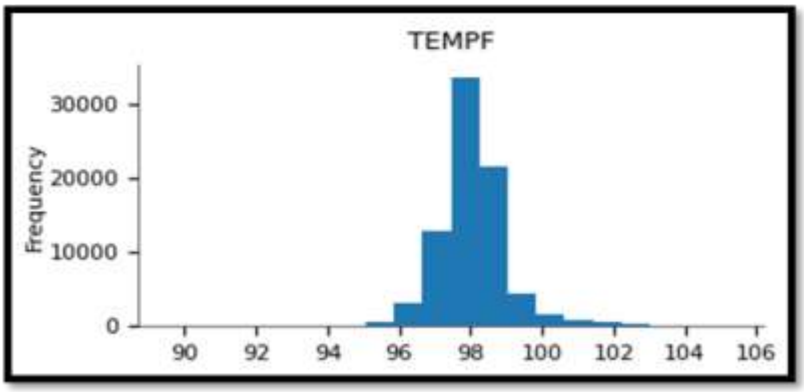
Table 1: Histogram, 2D KDE (Kernel Density Estimation) and Violin Plots for TEMPF, RESPR, PULSE, BPSYS, BPDIAS and POPCT w.r.t SCORE

Pramit Kumar Samant, Vinay Pathak & Waqar Ahmad (2026). Machine Learning Based Decision Support System on Electronic Health Record Data to Predict Health Score. International Journal of Multidisciplinary Research & Reviews, 5(7),239-274.

We have analyzed which attributes of the dataset are most important to obtain the health score value. Some visualization based on histogram analysis, two-dimensional kernel density estimations and violin plots are used to be shown table 1 respectively. A visual depiction of the distribution of data, a histogram shows how many occurrences there are in each bin i.e. RESPR, BPSYS and POPCT attributes measure the health score with respect to density to be shown in table 1. It may be applied to compare various datasets and find trends and patterns in the data (Zhang et al. 2023). A non-parametric approach for calculating the kernel density of a probability density function is called kernel density estimation. As a 2D kernel density plot, it is used to illustrate the normalized color density of a scatter plot. KDE plot is a graphical representation of the distribution of observed values in a data set, like the histogram. The highest values in the data are shown using a violin plot i.e. bivariant violin plot in table 1, a style of graph that combines a box plot with a density plot. It is a tool for displaying how numbers are dispersed. Violin plots provide summary numbers together with the frequency of each value, as opposed to just summary numbers. RESPR, BPSYS and POPCT attributes are important to distinguish the health scores from the different observations to be shown in table 1. The plot between health score and temperature (or other attributes) illustrates the relationship between these factors and overall health. Comparing health scores with different attributes, such as temperature, oxygen levels, etc., helps identify patterns or correlations that are critical for assessing patient health and predicting potential risks.

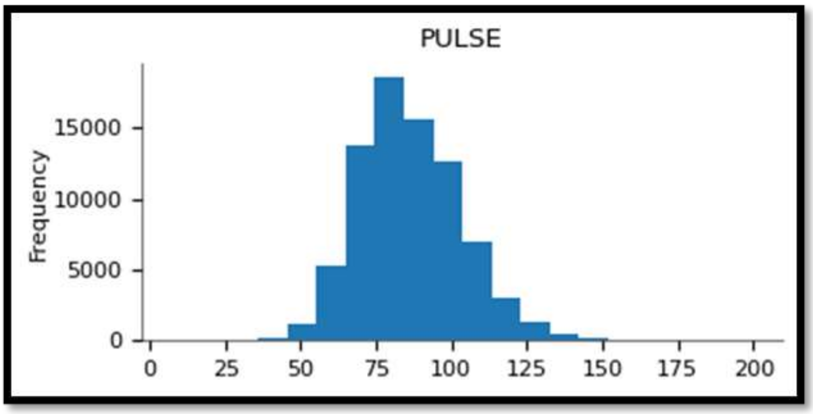
4.3 Frequency Distribution and Scatter plot

There is frequency distribution and scatter plot that is depicted all attributes do not depend on each other in tight bound fashion. When temperature, oxygen levels, and systolic pressure do not overlap in a plot, it means that these attributes display distinct and separate patterns. This indicates that each attribute independently contributes to understanding the health score, making them valuable for individual or combined analysis in the predictive model.

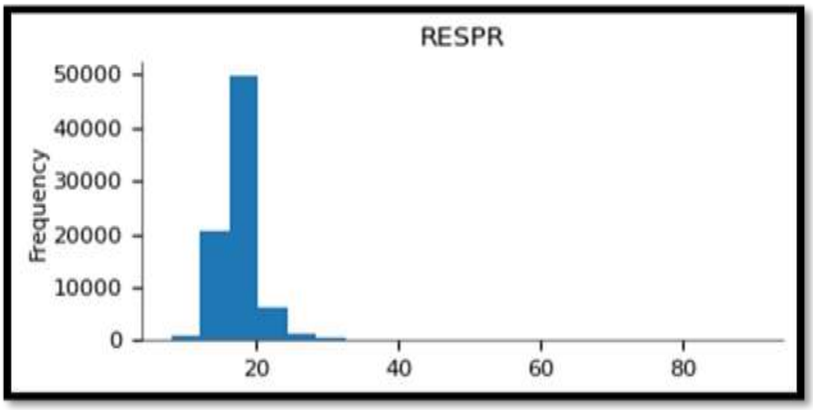


(a)

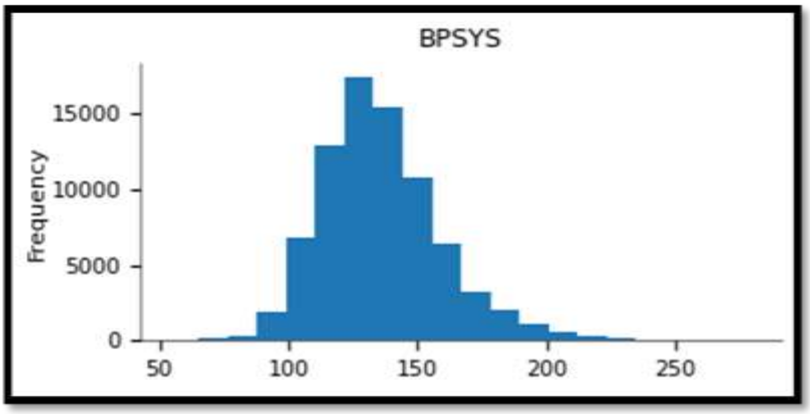
Pramit Kumar Samant, Vinay Pathak & Waqar Ahmad (2026). Machine Learning Based Decision Support System on Electronic Health Record Data to Predict Health Score. International Journal of Multidisciplinary Research & Reviews, 5(7),239-274.



(b)

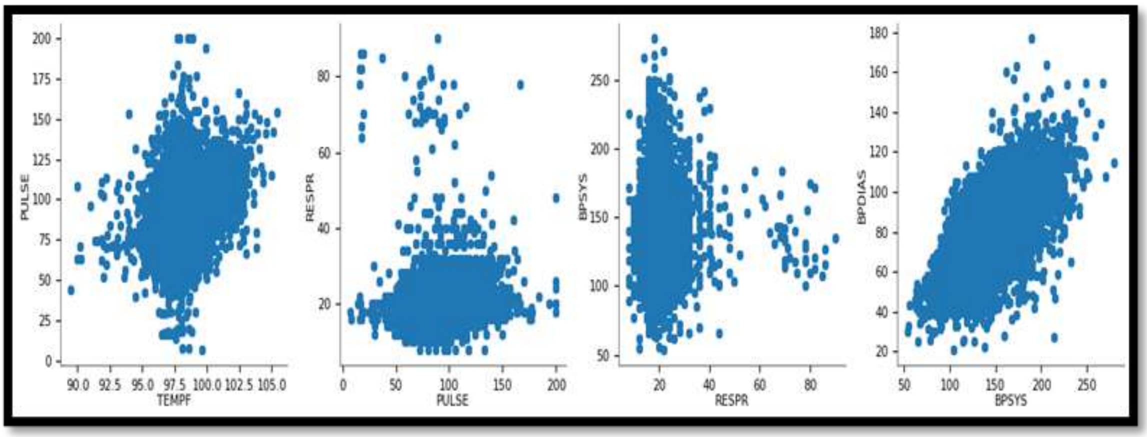


(c)



(d)

Pramit Kumar Samant, Vinay Pathak & Waqar Ahmad (2026). Machine Learning Based Decision Support System on Electronic Health Record Data to Predict Health Score. International Journal of Multidisciplinary Research & Reviews, 5(7),239-274.



(e)

Figure 11: (a), (b), (c), (d) are the frequency distributions of TEMPF, PULSE, RESPR AND BPSYS AND (e) scatter plot between (TEMPF and PULSE), (PULSE and RESPR), (RESPR and BPSYS) and (BPSYS and BPDIAS)

Frequency distribution with respect to each attribute i.e. temperature, pulse, respiration system and systolic blood pressure to be shown in figure 11.outlier and missing values not to be detected in frequency distributions of temperature, pulse, respiration system and systolic blood pressure as well as pattern of systolic blood pressure, pulse and respiration system to be same to be identified in figure 11 (a),(b),(c) and (d) respectively.

In scatter plot, independent variables like pulse, respiration system and systolic blood pressure are identified as same pattern with respect to dependent variables. However, noises are identified with dependent variables like RESPR, BPSYS and BPDIAS to be shown in figure 11(e). Therefore, RESPR, BPSYS and POPCT are important significance to distinguish health score.

4.4 Analysis of EHR dataset through SVM algorithm

```
Accuracy Score for Logistic Regression is :  71.3645126141983 %
Accuracy Score for Support Vector Machine is :  93.02656944095214 %
Accuracy Score for Random Forest is :  100.0 %
Accuracy Score for Gaussian Naives Bayes is :  74.29385634062527 %
Accuracy Score for K Nearest Neighbour is :  91.65199899421674 %
Accuracy Score for Linear Discriminant Analysis is :  71.82549660548152 %
Accuracy Score for Decision Tree is :  100.0 %
```

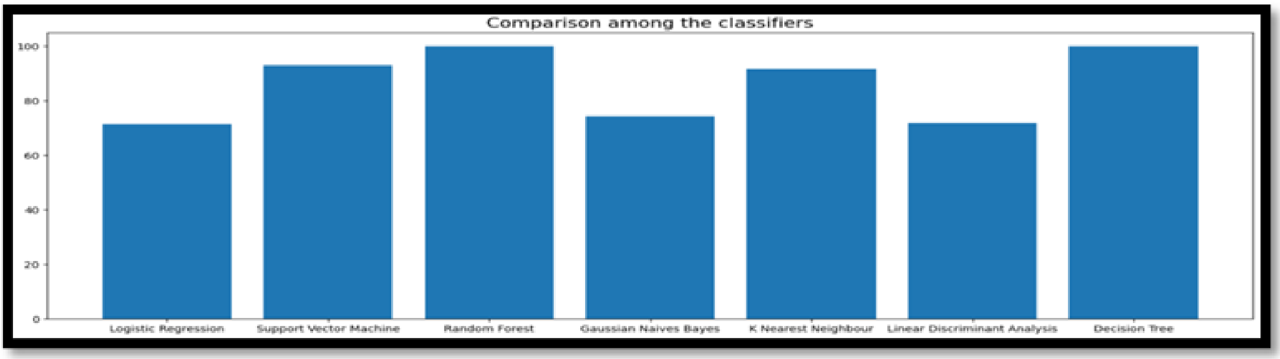
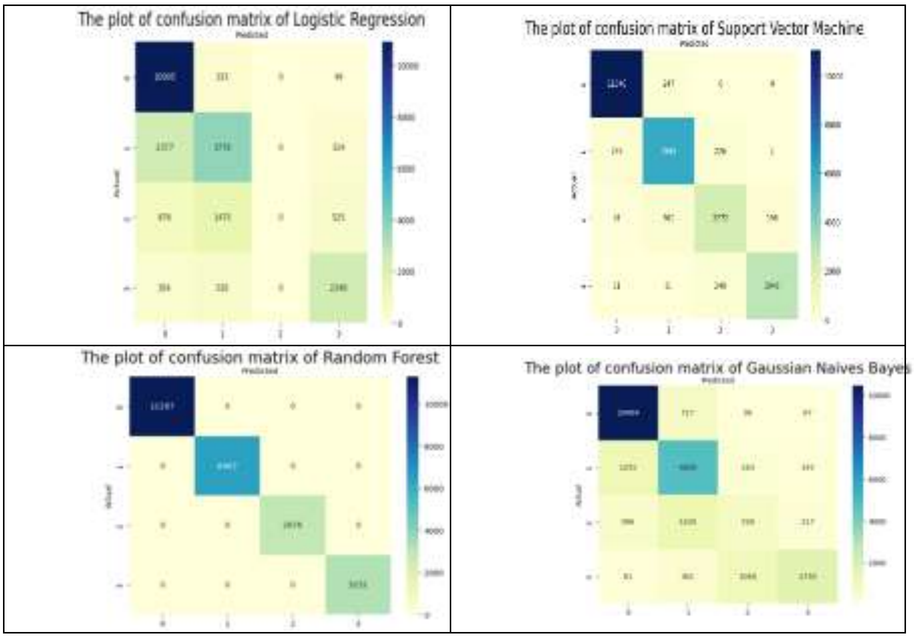


Figure 12: Comparison in accuracy among different Machine Learning Algorithms

From figure 12, SVM i.e., support vector machine has the highest classification's accuracy (near to 93%) among the logistic regression, random forest, Gaussian naves bayes, K nearest neighbor, Linear Discriminant Analysis and decision tree classifiers. The SVM (Support Vector Machine) algorithm calculates the health score using three parameters. These parameters were selected based on their significance to the health score, ensuring the model is both accurate and efficient.

4.4.1 Confusion Matrix

The concept of the multi class confusion matrix is like binary class matrix. The columns represent original class distribution, and the row represents predicated distribution by the classifiers. The plot and compare of confusion matrix of all the classifiers to be shown in figure 13.



ISSN (ELECTRONIC): 2945-3135
Pramit Kumar Samant, Vinay Pathak & Waqar Ahmad (2026). Machine Learning Based Decision Support System on Electronic Health Record Data to Predict Health Score. International Journal of Multidisciplinary Research & Reviews, 5(7),239-274.

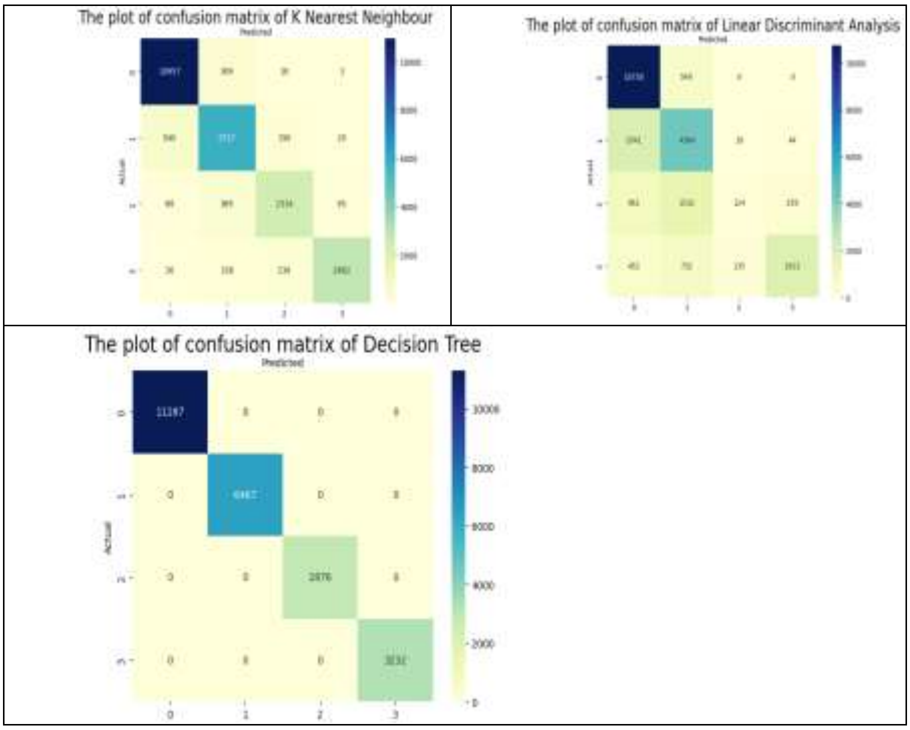


Figure 13: Comparing Confusion Matrix for All Models

The multiclass confusion matrix of SVM model to be mentioned in figure 13 has been converted into binary class confusion matrix to be shown in figure 14.

$$\begin{bmatrix} \mathbf{11040} & \mathbf{247} \\ \mathbf{352} & \mathbf{12223} \end{bmatrix}$$

Figure 14: Binary Class Confusion matrix for SVM Model

The value of true positive (TP) is 11040 that means sample belonging to positive class being classified correctly. And the value of true negative (TN) is 12223 that means sample belonging to negative class being classified correctly. Therefore, accuracy of SVM model more than 93%.

4.5 Analysis of EHR dataset through Artificial Nural Network

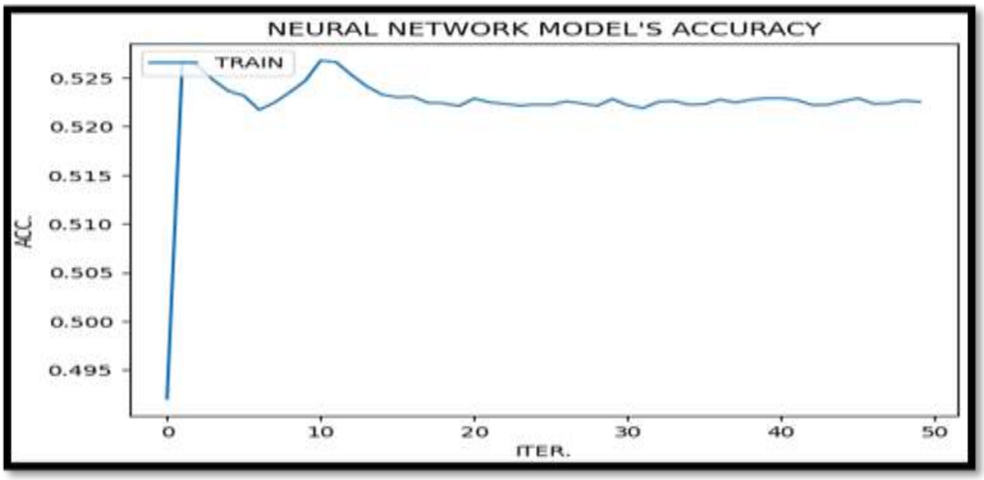
```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 dense (Dense)               (None, 12)                84

 dense_1 (Dense)             (None, 6)                 78

 dense_2 (Dense)             (None, 1)                 7

=================================================================
Total params: 169 (676.00 Byte)
Trainable params: 169 (676.00 Byte)
Non-trainable params: 0 (0.00 Byte)
_________________________________________________________________
```

Pramit Kumar Samant, Vinay Pathak & Waqar Ahmad (2026). Machine Learning Based Decision Support System on Electronic Health Record Data to Predict Health Score. International Journal of Multidisciplinary Research & Reviews, 5(7),239-274.

```
Epoch 50/50
5568/5568 [==============================] - 10s 2ms/step - loss: -191168112.0000
- accuracy: 0.5225
1740/1740 [==============================] - 3s 1ms/step - loss: -196886512.0000 -
accuracy: 0.5272
                           Accuracy Measurement: 52.72
```

(After 50 iterations accuracy of this sequential NN model have reached at 52.72%)

Figure 15: Sequential Model Layered Format (Deep Learning mechanism)



(Train Model accuracy graph for neural network)

4.6 New sample testing by known Model prediction

The values are shown in normal range

What is your body RESPR: 12 to 18=18

The values are shown in normal range

What is your body BPSYS: 120 to 129=121

The values are shown in normal range

What is your POPCT: 95 to 100=97

New Data

[18,121,97]

Pramit Kumar Samant, Vinay Pathak & Waqar Ahmad (2026). Machine Learning Based Decision Support System on Electronic Health Record Data to Predict Health Score. International Journal of Multidisciplinary Research & Reviews, 5(7),239-274.

```
Classification and Accuracy report for Logistic Regression
Label for HealthScore =  [0]
----------------------------------------------------------
Classification and Accuracy report for Support Vector Machine
Label for HealthScore =  [0]
----------------------------------------------------------
Classification and Accuracy report for Random Forest
Label for HealthScore =  [0]
----------------------------------------------------------
Classification and Accuracy report for Gaussian Naives Bayes
Label for HealthScore =  [2]
----------------------------------------------------------
Classification and Accuracy report for K Nearest Neighbour
Label for HealthScore =  [0]
----------------------------------------------------------
Classification and Accuracy report for Linear Discriminant Analysis
Label for HealthScore =  [0]
----------------------------------------------------------
Classification and Accuracy report for Decision Tree
Label for HealthScore =  [0]
----------------------------------------------------------
```

Prediction by Deep Learning Algorithm

```
[ ]  y_pred = model_NN.predict(colX_test[[colX.shape[0]]])
     print("Label for HealthScore = ", y_pred)

     1/1 [==============================] - 0s 20ms/step
     Label for HealthScore =  [[0.]]
```

We have shown that prediction by both approach machine learning algorithms as well as deep learning algorithms reveals that this dataset was satisfied as in well. In our dataset analysis, SVM is the most effective classifier than another ML classifier and simple sequential neural network model to be shown in figure 12 and figure 14. Therefore, only three important attributes have been made to focus on the most relevant features that have the highest impact on the health score. This simplifies the model, making it more efficient and less prone to overfitting, while still maintaining a high level of accuracy.

5. Conclusion

The focus of e-health has mostly been on providing detailed technical responses to inquiries without taking knowledge acquisition into account. Therefore, only three important attributes have been made to focus on the most relevant features that have the highest impact on the health score. This simplifies the model, making it more efficient and less prone to overfitting, while still maintaining a high level of accuracy. The findings indicate that information and communication technology may unquestionably aid in the sharing of knowledge in clinics. Making ensuring that the information acquired is accurate and dependable while handling and organizing enormous volumes of data

Pramit Kumar Samant, Vinay Pathak & Waqar Ahmad (2026). Machine Learning Based Decision Support System on Electronic Health Record Data to Predict Health Score. International Journal of Multidisciplinary Research & Reviews, 5(7),239-274.

properly is crucial. The concept of e-health is useful since it offers guidance on how to take care of oneself independently. Online, people discuss health issues frequently, but they seldom divulge their knowledge of these issues on social media. This study aims to investigate the impact of knowledge on e-health. The future scope of the research could include, expanding the dataset by incorporating more diverse data from additional sources. Refining the model by experimenting with more advanced machine learning algorithms. Applying the model in clinical settings for real-world validation. Incorporating additional physiological parameters to further improve the health score's accuracy.

CRediT authorship contribution statement

Pramit Kumar Samant: Writing – original draft, Methodology. **Vinay Pathak**: Writing – original draft, Methodology. **Waqar Ahmad**: Conceptualization, Writing, Formal analysis, Investigation, Software, Validation.

Data availability: The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest: The authors declare that they have no competing interests.

Funding Declaration: We declare that this manuscript received no funding from any sources.

AUTHOR(S) CONTRIBUTION

The writers affirm that they have no connections to, or engagement with, any group or body that provides financial or non-financial assistance for the topics or resources covered in this manuscript.

CONFLICTS OF INTEREST

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

PLAGIARISM POLICY

All authors declare that any kind of violation of plagiarism, copyright and ethical matters will take care by all authors. Journal and editors are not liable for aforesaid matters.

SOURCES OF FUNDING

The authors received no financial aid to support for the research.

REFERENCES

[1]. Iman Shafi,"A Comprehensive Review of Recent Advances in Artificial Intelligence for Dentistry E-Health,"in journal, diagnostics., vol. 13, issue13,2023,pp.1–30.

Pramit Kumar Samant, Vinay Pathak & Waqar Ahmad (2026). Machine Learning Based Decision Support System on Electronic Health Record Data to Predict Health Score. International Journal of Multidisciplinary Research & Reviews, 5(7),239-274.

[2] S. Sengan, "Medical Information Retrieval Systems for E-Health Care Records Using Fuzzy-Based Machine Learning Model".Belmont, CA, USA:Wadsworth, 2020, pp. 123–135.

[3] A.Ahmed, "Applicability of Portable Health Clinic to Ageing Society," DAPI 2014. Lecture Notes in Computer Science, vol 8530. Springer, Cham. https://doi.org/10.1007/978-3-319-07788-8_49.

[4]. S. J. Rigatti, "Random Forest," January 2017 Journal of Insurance Medicine 47(1):31-39 DOI:10.17849/insm-47-01-31-39.

[5]. A Tolera, "Electronic medical record use and associated factors among healthcare professionals at public health facilities in Dire Dawa, eastern Ethiopia: A mixed-method study," Front. Digit. Health, 23 August 2022 Sec. Health Informatics Volume 4 – 2022, https://doi.org/10.3389/fdgth.2022.935945.

[6]. Bonner, C.,"Implementing cardiovascular disease prevention guidelines to translate evidence-based medicine and shared decision making into general practice: theory-based intervention development, qualitative piloting and quantitative feasibility," Implementation Science, 14(1), pp.1-14, 2019.

[7]. Coulthard, P.,"Dentistry and coronavirus (COVID-19)-moral decision-making." British Dental Journal, 228(7), pp.503-505, 2020.

[8]. Mansel, B. and Einion, A., "It's the relationship you develop with them': emotional intelligence in nurse leadership. A qualitative study", British Journal of Nursing, 28(21), pp.1400-1408, 2019.

[9]. Satamraju, K.P. and Balakrishnan, M., "A secured healthcare model for sensor data sharing with integrated emotional intelligence" IEEE Sensors Journal, 22(16), pp.16306-16313, 2022.

[10]. Campolina, A.G., Suzumura, E.A., Hong, Q.N. and de Soarez, P.C.," Multicriteria decision analysis in health care decision in oncology: a systematic review. Expert review of pharmacoeconomics & outcomes research", 22(3), pp.365-380, 2022.

[11]. Yousif, M.G.," Advancements in Medical Research in Iraq: A Comprehensive Review of Emerging Insights", Medical Advances and Innovations Journal, 1(2), p.2,2022.

[12]. Bouh, M.M., Hossain, F. and Ahmed, A.," A Machine Learning Approach to Digitize Medical History and Archive in a Standard Format", 2023.

[13]. J Dugué, M., Sirost, O. and Dosseville, F., "A literature review of emotional intelligence and nursing education", Nurse Education in Practice, 54, p.103124,2021.

Pramit Kumar Samant, Vinay Pathak & Waqar Ahmad (2026). Machine Learning Based Decision Support System on Electronic Health Record Data to Predict Health Score. International Journal of Multidisciplinary Research & Reviews, 5(7),239-274.

[14]. Rai, R., Vats, R., Vermas, G. and Kumar, M.," Enhancing medical field with e-health and AI: A systematic review", In AIP Conference Proceedings (Vol. 2721, No. 1). AIP Publishing, 2023, July.

[15]. Bell, M. and Sheridan, A.,"How organisational commitment influences nurses' intention to stay in nursing throughout their career", International Journal of Nursing Studies Advances, 2, p.100007,2020.

[16]. Vatandoost, M. and Litkouhi, S.,"The future of healthcare facilities: how technology and medical advances may shape hospitals of the future", Hospital Practices and Research, 4(1), pp.1-11, 2019.

[17]. Castaneda, J., Calvet, L., Benito, S., Tondar, A. and Juan, A.A., "Data science, analytics and artificial intelligence in e-health: trends, applications and challenges", SORT-Statistics and Operations Research Transactions, pp.81-128, 2023.

[18]. Arasanmi, C.N. and Krishna, A.,"Employer branding: perceived organisational support and employee retention–the mediating role of organisational commitment. Industrial and Commercial Training", 51(3), pp.174-183,2019.

[19]. Godi, B., Viswanadham, S., Muttipati, A.S., Samantray, O.P. and Gadiraju, S.R., "March. E-healthcare monitoring system using IoT with machine learning approaches", in 2020 international conference on computer science, engineering and applications (ICCSEA) (pp. 1-5). IEEE.

[20]. Mohapatra, S.K. and Mohanty, M.N., "Big data classification with iot-based application for e-health care", in Cognitive big data intelligence with a metaheuristic approach (pp. 147-172). Academic Press, 2022.

[21]. Kumar, S., Bhatt, R. and Ganguly, D.G.,"Organizational behavior", Academic Guru Publishing House,2022.

[22]. McKenna, E.,"Business psychology and organizational behavior", Routledge, 2020.

[23]. Weber, W.G., Unterrainer, C. and Höge, T., "Psychological research on organisational democracy: A meta-analysis of individual, organisational, and societal outcomes", Applied Psychology, 69(3), pp.1009-1071, 2020.

[24]. Mousa, M., Massoud, H.K. and Ayoubi, R.M.," Gender, diversity management perceptions, workplace happiness and organisational citizenship behaviour. Employee Relations", The International Journal, 42(6), pp.1249-1269, 2020.

Pramit Kumar Samant, Vinay Pathak & Waqar Ahmad (2026). Machine Learning Based Decision Support System on Electronic Health Record Data to Predict Health Score. International Journal of Multidisciplinary Research & Reviews, 5(7),239-274.

[25]. Zhang, Y., Xu, P., Sun, Q., Baral, S., Xi, L. and Wang, D.," Factors influencing the e-health literacy in cancer patients: a systematic review", Journal of Cancer Survivorship, 17(2), pp.425-440, 2023.

[26]. Cavaness, K., Picchioni, A. and Fleshman, J.W.,"Linking emotional intelligence to successful health care leadership: the big five Model of Personality", Clinics in colon and rectal surgery, 33(04), pp.195-203., 2020.

[27]. Halian, A., Halian, I., Burlakova, I., Shevchenko, R., Lappo, V., Zhigarenko, I. and Popovych, I., "Emotional intelligence in the structure of adaptation process of future healthcare professionals. Revista Inclusiones, pp.447-460, 2020.

[28] AMarcin Michał Mirończuk, Jaroslaw Protasiewicz,"A Recent Overview of the State-of-the-Art Elements of Text Classification", Expert Systems with Applications 106,DOI:10.1016/j.eswa.2018.03.058,2018.

[29] Jinhyuk Lee, Wonjin Yoon, "BioBERT: a pre-trained biomedical language representation model for biomedical text mining", v4,18 Oct 2019.

[30] VYash Pawar,Aron Henriksson,"Leveraging Clinical BERT in Multimodal Mortality Prediction Models for COVID-19",IEEE 35th International Symposium on Computer-Based Medical Systems, 2022.

[31] Laila Rasmsy,Yang Xiang,"Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction",npj Digit. Med. 4,20 May 2021.

[32] Jagadamba G,Shashidhar R,Gururaj H L,Vinayakumar Ravi, Meshari Almeshari,"Electronic Health Record (EHR) System Development for Study on EHR Data-based Early Prediction of Diabetes Using Machine Learning Algorithms" In DOI: 10.2174/18750362-v16-e230906-2023-15

[33] Nuria C. Machine learning applied to diagnosis of human diseases: A systematic review. Appl Sci MDPI Acad Comm 2020; 10(15): 5135.

[34] Asan O, Henry NY, Betty C, *et al.* How physician electronic health record screen sharing affects patient and doctor non-verbal communication in primary care. Patient Educ Couns 2015; 98(3): 310-6.

[35] Diabetes. Available from: https://www.kaggle.com/johndasilva/diabetes

[36] Cabana M D, Rand C S, Powe N R, et al. Why don't physicians follow clinical practice guidelines? a framework for improvement. JAMA 1999; 282(15): 1458-65. [http://dx.doi.org/10.1203/00006450-199904020-00719]

Pramit Kumar Samant, Vinay Pathak & Waqar Ahmad (2026). Machine Learning Based Decision Support System on Electronic Health Record Data to Predict Health Score. International Journal of Multidisciplinary Research & Reviews, 5(7),239-274.

[37] James R. The American College of surgeons, minimum standards for hospitals, and the provision of high-quality laboratory services. Arch Pathol Lab Med 2017; 141(5): 704-17.

[38] Stausberg J, Koch D, Ingenerf J, Betzler M. Comparing paper-based with electronic patient records: Lessons learned during a study on diagnosis and procedure codes. J Am Med Inform Assoc 2003; 10(5): 470-7. [http://dx.doi.org/10.1197/jamia.M1290]

[39] Tang PC, McDonald CJ. Electronic health record systems. In: Shortliffe EH, Cimino JJ, Eds. Biomedical Informatics Health Informatics. New York, NY: Springer 2006. [http://dx.doi.org/10.1007/0-387-36278-9_12]

[40] Elyse O, Steve EA, Alan RS, et al. Clinical decision support for recognition and management of hypertension: a randomized trial. Pediatrics 2018; 141(2): e20172954.

[41] William R. The electronic medical record: Promises and problems. J Am Soc Inf Sci 1995; 46(10): 772-6.

[42] Eta SB. Clinical decision support systems. Health Informatics book series, Springer Science+ Business Media,LLC 2007; 233: 3-22.

[43] Davis D A. Translating guidelines into practice: A systematic reviewof theoretic concepts, practical experience and research evidence in the adoption of clinical practice guideline. CMAJ 1997; 157(4): 408-16.

[44] Kwok R, Dinh M, Dinh D, Chu M. Improving adherence to asthma clinical guidelines and discharge documentation from emergency departments: Implementation of a dynamic and integrated electronic decision support system. Emerg Med Australas 2009; 21(1): 31-7.[http://dx.doi.org/10.1111/j.1742723.2008.01149.x]

[45] Jagadamba G, Krishna ELS, Amogh JP, Abhishek BB, Manoj HN. Adaptive electronic health records management and secure distribution using blockchain. In: Lecture Notes in Networks and Systems. Singapore: Springer 2022; 481.

[46] Pai, M.M.M., Ganiga, R., Pai, R.M. et al. Standard electronic health record (EHR) framework for Indian healthcare system. Health Serv Outcomes Res Method 21, 339–362 (2021). https://doi.org/10.1007/s10742-020-00238-0

[47] Avendaño-Valencia, L. D., & Fassois, S. D. (2015, July). Natural vibration response-based damage detection for an operating wind turbine via random coefficient linear parameter varying AR modelling. In Journal of Physics: Conference Series (Vol. 628, No. 1, p. 012073). IOP Publishing.

[48] Abedi, M., Norouzi, G. H., & Bahroudi, A. (2012). Support vector machine for multi-classification of mineral prospectivity areas. Computers & Geosciences, 46, 272-283.

[49] Sluiter, R., & Pebesma, E. J. (2010). Comparing techniques for vegetation classification using multi-and hyperspectral images and ancillary environmental data. International Journal of Remote Sensing, 31(23), 6143-6161.

[50] Dietrich, R., Opper, M., & Sompolinsky, H. (1999). Statistical mechanics of support vector networks. Physical review letters, 82(14), 2975.

[51] Suykens, J. A., & Vandewalle, J. (1999). Least squares support vector machine classifiers. Neural processing letters, 9, 293-300.

[52] Shen, T., Li, H., & Huang, X. (2010). Active volume models for medical image segmentation. IEEE transactions on medical imaging, 30(3), 774-791.

[53] Tsai, C. F., Hsu, Y. F., Lin, C. Y., & Lin, W. Y. (2009). Intrusion detection by machine learning: A review. expert systems with applications, 36(10), 11994-12000.

[54] Smeraldi, F., & Bigun, J. (2002). Retinal vision applied to facial features detection and face authentication. Pattern recognition letters, 23(4), 463-475.