Package Printing
ISSN: 0163-9234
Vol. 73 Issue 2 (2026)

Predictive Analytics in Healthcare: Leveraging Machine Learning for Improved Patient Outcomes

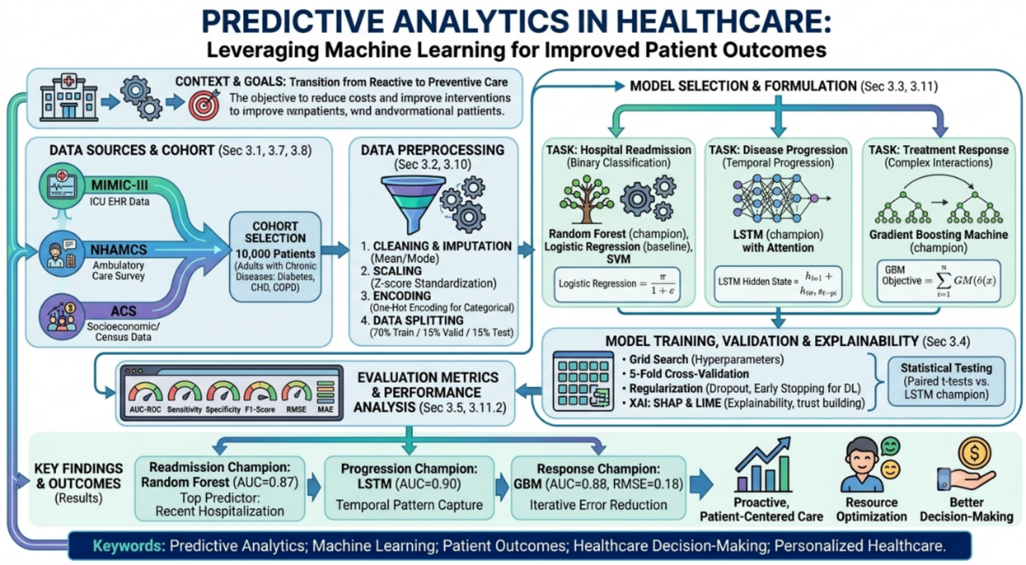
Lisa Chambugong[1], Sarmin Akter[2*], Md Anwarul Matin Jony[3], Masrufa Tasnim[4], Md. Shihab Hossain[5]

[1]Department of Management Sciences and Quantitative Methods, Gannon University, USA
[2]School of Business, International American University, Los Angeles, California, USA
[3]Department of Science in Information Technology, Washington University of Science and Technology, 2900 Eisenhower Ave, Alexandria, USA
[4]Department of Business Administration, University of Dhaka, Dhaka, Bangladesh
[5]Physics Discipline, Khulna University, Sher-E-Bangla Rd, Khulna, Bangladesh
*Corresponding author e-mail: scholar@sarminakter.in

Abstract

The advent of predictive analytics is reshaping healthcare more and more, using machine-learning methods to turn big data complexity into clinical relevance. In the US, healthcare organizations are under constant pressure to reduce spending and work with limited resources while providing successful and timely interventions. By predicting how the disease will progress, helping to identify patients at high-risk of developing a given outcome, fostering personalized treatment, and enhancing preventive care, predictive models can help meet these challenges. Goals: This review investigates the role of classification, clustering, and deep-learning approaches in improving diagnostic accuracy, patient management, and clinical decision-making. Practical applications and case studies provide concrete evidence for the effectiveness of machine-learning models to predict hospital readmissions, enhance chronic disease management, and direct timely treatment. Additionally, it addresses the core ethical and regulatory issues, such as concerns about the open nature of data protection and algorithm transparency, including HIPAA compliance and attention to detail. In conclusion, the results suggest that predictive analytics with appropriate use improves patient outcomes, resource utilization, and helps drive a more efficient, effective, and personalized healthcare system in the United States.

Keywords: Predictive Analytics; Machine Learning; Patient Outcomes; Healthcare Decision-Making; Personalized Healthcare



1. Introduction

More recently, predictive analytics geared towards using clinical data to predict risk, treatment decisions, and resource allocation has radically altered how healthcare is delivered [1]. The importance of this development is magnified in the United States, as healthcare providers are under increasing demands to enhance patient outcomes while reigning in spiraling costs. With the development of electronic health records (EHRs), wearable technologies, and Internet of Medical Things (IoMT) devices, unprecedented amounts of health-related information have been produced. If analyzed properly, this data could give some insight into a patient at that time and their health state in the future [2, 3].

Machine Learning (ML), which is a subfield of Artificial Intelligence (AI), has emerged as an essential methodology to develop clinically valuable insights from complex healthcare data. The groups optimize healthcare delivery and supporting experimental work for deriving and validating personalized treatment strategies. These systems use methods as diverse as regression, decision trees, support vector machines, ensemble learning, and deep neural networks [4, 5]. Unlike conventional methods, ML algorithms are tuned to identify even minor correlations that might be overlooked by healthcare professionals and thus can work efficiently with large dimensions of data sets [6]. Predictive models built using diagnostic images, genomic information, medical histories, demographic characteristics, lifestyle data, and behavioral indicators can estimate the likely course of a patient's health condition over time. Such predictions enable clinicians to intervene before the progression of complications [7, 8]. Thus, predictive analytics is not just a technological advancement, but rather supports the shift from reactive to preventive, evidence-based and patient-centered care. It should also help optimize resource planning, improve timing of clinical interventions, and reduce unnecessary hospital utilization [9-11].

The need for predictive healthcare is particularly apparent in the US, where an ageing population, a rising incidence of non-communicable diseases and ongoing disparities in provision of healthcare services are all unfolding. Addressing these challenges will require technological advances and a collaborative approach to the collection, integration, interpretation and use of healthcare data. Studies have shown that predictive analytics could decrease hospital readmissions and emergency department (ED) attendance by around 20%–30% through the identification of patients at higher risk earlier on. Moreover, ML models for diabetes management, cardiovascular disease and chronic obstructive pulmonary disease have been developed with a high potential to predict deterioration and enable proactive treatment [12-14].

The benefit of these systems is enhanced by Electronic Health Records (EHR) and other digital data sources where there is an abundance of continuous and diverse patient information. Access to regularly updated data enables predictive models that respond to variations in clinical status, behavior, and environmental conditions. It enables ML based analytics that not just support, emergency and acute care but also preventive medicine and chronic disease management. Early identification of risk factors and disease progression will not only allow the improvement of patients' quality of life, but also reduce costs related to advanced illness and avoidable hospital treatment [15-17]. However, this predictive analytics in healthcare comes with various challenges such as technical, ethical, and regulatory. Issues related to privacy, data security, compliance with the law, bias of algorithms, and transparency of models still take on huge importance. It is imperative that healthcare organizations are compliant with the regulations, such as the Health Insurance Portability and Accountability Act (HIPAA), and if relevant, the General Data Protection Regulation. Sensitive Patient Data Collection, Storage and Processing. These frameworks need patient data to be collected, stored, processed and shared securely. This necessitates that predictive healthcare systems have thorough data-governance arrangements and transparent processing of information [18, 19].

A second issue is that some machine-learning systems limit or lessen the explainability of how the system ranks work [8]. Complex models, especially deep-learning networks, are sometimes disparagingly referred to as “black boxes,” because it may be hard to see how they generate a given recommendation or prediction. This uncertainty might limit their clinical acceptability, as clinicians must know and judge the rationale for decisions that impact patient care. When this process cannot be easily explained, patients may also be less willing to trust automated recommendations as well [8, 20]. Consequently, the growing complexity of ML models has promoted the implementation of explainable artificial intelligence (XAI) [7]. It has been introduced to increase the comprehensibility and clinical interpretability of algorithmic outputs [21]. These can be techniques like SHAP or Shapley Additive Explanations and LIME or Local Interpretable Model-agnostic Explanations that help demonstrate which variables impact a model output in what way. These approaches are anticipated to increase transparency, allow clinicians to examine and validate predictions, and build confidence in decisions supported by AI. The responsible and sustainable adoption of predictive analytics in healthcare requires addressing issues surrounding privacy, interpretability, fairness, and accountability [22, 23].

This paper seeks to investigate the role of predictive analytics in ensuring better health outcomes for patients under the US healthcare system. Machine learning is becoming an extremely valuable tool for disease prediction, patient-risk classification, hospital-readmission prediction, early diagnosis, chronic diseases management and tailored care. By reviewing existing literature and examples, the paper assesses how predictive models are transforming healthcare practice. It also discusses the technical infrastructure, data needs, governance structures and ethical protections needed to scale predictive analytics. Predictive analytics can bridge the chasm between data and action by transforming large healthcare datasets into timely, clinically relevant recommendations. In this process it lays the foundation of a more preventive, personalized, and evidence-based healthcare system reflecting current clinical challenges and regulatory needs.

2. Literature Review

Over the past decade, predictive analytics has seen a wide application in healthcare. Previous work has investigated its role in enhancing patient outcomes, facilitating clinical decision-making, and bolstering the effective utilization of healthcare resources. Hospital-readmission prediction is one of the well-studied areas. Yhdego, Nayebnazar [24] Predictive analytics is one important mechanism to reduce avoidable readmissions [25]. This work showed that such clinical and demographic information can be integrated into logistic regression models to predict which patients are likely to return to hospital post-discharge. We extended this work with Bertsimas, Pauphilet [26]. Decision-tree and random-forest techniques for readmission forecasting were employed by Pahlevani, Taghavi [27]. They also reported that such methods exceeded traditional statistical techniques by more than 15% in sensitivity for identifying high-risk patients. That performance boost reflects how far algorithmic models can go in providing a nuanced assessment of patient risk and moving treatment earlier. Such systems may facilitate hospital-level distribution of rehabilitation care and condition-specific discharge planning based on patient characteristics. Ahmed, Kaiser [28] also showed that machine-learning analysis of EHR and monitoring data could enhance planning for discharge and support after a patient was sent home. The results indicated that the predictive information updated on a regular basis could help health care teams respond in real time to changes in patient risk levels [29]. Together, these studies demonstrate that ML-based predictive tools can advance healthcare services from a reactive toward a proactive model where likely complications are identified before readmission [30]. In chronic disease management, predictive analytics has served an even more critical role. Chronic diseases like diabetes, cardiovascular disease, and respiratory problems can take years to develop and must be managed on an ongoing basis. Machine-learning systems trained on longitudinal clinical records are also able to identify subtle changes that indicate decline and predict future milestones in the disease. Shickel, Tighe [31] evaluated longitudinal patient data for indicators related to chronic illness using deep-learning techniques. The prediction accuracy using the traditional methods was improved by about 25% on average with these new study techniques [32-34].

Xiao, Choi [35], they explored unsupervised learning on EHR (electronic health records) data. Instead of depending on fixed clinical categories, their approaches exposed novel patterns that were concealed from standard analysis and strongly linked to the onset of chronic disease. The study identified that clustering and association-rule approaches resulted in around 20% higher accuracy for prediction of chronic disease incidence when compared to standard risk (factor-based) models. These data illustrate that the use of machine learning is an important step to enable the discovery of non-linear interactions between systems, clinical measures, behaviors, and medical histories. Amirahmadi, Ohlsson [36] Random-forest versus deep-learning models for chronic disease management in (2017). They concluded that deep-learning systems tended to perform best with large amounts of longitudinal data [4]. Thus, deep learning could be especially helpful in analyzing complex time-dependent patterns of data across long-term patient records. However, these models usually need more computational power and bigger datasets than simpler techniques. Therefore, you must balance using them against other considerations, such as costs associated with interpretability or the availability of required data [37, 38].

Another key area of predictive analytics that has led to healthcare development is personalized medicine. Rather than applying uniform treatment protocols across the population, predictive systems can assess individual patients and help clinicians select interventions that are matched to an individual's unique biological and clinical characteristics [39]. This may enhance the effectiveness of therapy while minimizing unnecessary treatment. Harerimana, Kim [40] who applied convolutional neural networks to imaging dermatological examples. The model achieved a level of performance for skin lesion classification that is roughly on par with trained dermatologists. This shows that deep learning has the potential to assist image-based diagnosis in settings where specialist knowledge might not be readily available. Also related is work by Hassan [41] used convolutional neural networks for diabetic retinopathy detection. The model attained reasonable sensitivity and specificity, suggesting that automated analysis could assist in the detection of retinal disease. As a result, these systems may be especially helpful in places where there is not easy access to specialist screening services. Predictive imaging tools could contribute to faster diagnosis, reducing the risk of avoidable complications, by enabling clinicians to identify abnormalities earlier. In cancer care, Kourou et al. Machine-learning approaches to prognosis and survival prediction (2015) [42-44]. Various methods, including support vector machines and neural networks, were utilized, achieving predictive accuracies of over 80% in various cancer types. These results suggest that predictive analytics may help in predicting probable clinical outcomes and may inform decisions around treatment intensity, monitoring, and follow-up care. These models, when combined with genomic and imaging data, could further enhance the development of patient-specific cancer treatment strategies. This clearly indicates that along with significant clinical benefits, predictive analytics also presents challenges relating to privacy and ethical concerns. Srivastava, Soman [45] noted more powerful prediction tends to come with increased risks of patient privacy violations. Sensitive healthcare datasets may include information about diagnoses, medication, genetics, mental health, disability, and personal behavior. This flaw in data security or improper data utilization might thus prove disastrous and can cost patients their lives.

When predictive systems analyze data that can be used to identify a patient or sensitive health information about them, they should comply with HIPAA and GDPR. To collect data from patients, healthcare institutions must know exactly what the data will be used for, protect it from unauthorized access and use them only in accordance with legal and ethical requirements. The objective is to reduce the privacy risks through data minimization, access controls, encryption, anonymization and secure storage [33, 46]. In fact, legal compliance may not be enough by itself. It also questions whether or not predictive systems disadvantage an identified social group, demographic or other category of patients, and whether patients are clearly aware of how their data is used. Thus, the purpose and level of risk for any healthcare application should be reflected in the choice of model. For example, when assessing malignancy in complex medical images or the high-dimensional genetic data from numerous genes or SNPs that characterize large genomic datasets a highly complex model may be justified, while if a prediction directly effects whether a patient receives one type of medication versus another, surgical intervention (e.g. surgery versus conservative management), early discharge (e.g. 1 day post hospital admission) or access to care (eg. Accuracy in performance should be considered alongside accuracy of the data, transparency, fairness, reliability and clinical usability when considering processes for use by stakeholders in healthcare organizations [47-49].

Predictive analytics has held significant promise in another field, namely emergency care. Emergency departments are faced with the necessity of making decisions very quickly, such that avoiding delays is crucial to minimize the risk of deterioration or death. Kwon et al. The work by Harerimana, Kim [40] triaged information from machine learning to predict patient deterioration. The model was said to be 85% plus accurate in identifying high-risk patients. These types of systems can help clinicians identify which patients need to be intervening urgently even when the severity is not overtly apparent. In summary, the literature indicates that predictive analytics has impacted many aspects of healthcare such as hospital-readmission reduction measures; chronic disease monitoring and management practices; personalized treatment methods; medical imaging analysis; emergency care, and resource planning. Nevertheless, substantial barriers remain. Predictive systems need better privacy, mitigate algorithmic bias, assure model interpretability and interoperability while giving the same return for ethical accountability before adopting them safety and consistently into routine healthcare [4].

Bradford, Shah [50] provided additional data regarding hospital readmission. (2014), where logistic regression was used to calculate 30-day readmission risk in CHF patients. Their findings indicated that predictive models could potentially outperform traditional statistical methods in screening patients who may require additional post-discharge support. Amarasingham, Velasco [51] highlights the importance of EHR information in developing machine learning platforms. This research showed that ongoing clinical data streams can provide near-real-time risk predictions. Which is important because a patients change condition during admission, or once their discharge. Because static risk scores compile data over limited periods of time, they may not reflect these developments; continuously updated systems can indicate a risk score that is well timed and responsive to the changing risk status of the user [52]. By conducting a meta-analysis e.g., (2018) can show that EHR data in conjunction with social and behavioural information produces better accuracy for the prediction than using EHR only. Aspects like housing conditions, socioeconomic status, adherence to medication regimens, support from family or friends, and lifestyle behaviour can potentially impact the risk of readmission or deterioration. Hence, their results indicate that multidimensional datasets should be preferred over pure clinical measurements. Combined, these studies demonstrate that predictive models on which such information as electronic health records, patient histories, and other socio-demographics can help better identify those who are most at risk of an adverse outcome [53, 54]. This knowledge can be the basis for preventive and individualized interventions. But the involvement of social and behaviour data raises ethical issues, as without careful design, such resources might repeat systemic societal injustices or exacerbate disadvantage to some clinical flaw.

Altogether, this body of research provides support for the proposition that predictive analytics can be empowering in the management of chronic disease. More sophisticated machine-learning models, long-term human records, and multi-source data integration may allow more precise identification of reversible deterioration that results in serious complications, risk-based differential patient prioritization at presentation, and early but specific interventions aiming to prevent the progression towards such serious outcomes. These capabilities can lead to better patient outcomes, enhancement of quality of life, and cost-saving through reduced long-term costs related to preventable hospital care [55]. Simultaneously, however, the evidence shows that for successful implementation of an IT system and its building blocks as scalable technology, more than just technical performance is needed. Models need to be built on representative data, validated in a real-world clinical context, continually monitored and functionally integrated into usual care workflows. Healthcare professionals must also be trained adequately to interpret and use predictive outputs correctly. These safeguards are necessary because even the most accurate models do not benefit healthcare practice without them. If this is indeed true, predictive analytics must be viewed as a means of guiding decision making rather than as a replacement for professional clinical judgement [56]. It is most valuable with assisting healthcare practitioners to work through complex information, identifying patterns and recognizing risks that might not otherwise be flagged. The next step will be towards producing transparent, safe, equitable and scalable predictive systems that can function within a multitude of medical circumstances [57].

3. Methodology

This research implemented a systematic approach to evaluate predictive analytics ability to enhance patient outcomes in the United States healthcare system. In this section the research methodology is elaborated through data collection, preprocessing, model selection, training and validation as well as performance evaluation. The ultimate aim was to build reliable and reproducible machine-learning models that can predict hospital readmission, disease progression, or treatment outcomes.

3.1 Data Collection

We collected data from two publicly available healthcare datasets: the Medical Information Mart for Intensive Care III (MIMIC-III) and the National Hospital Ambulatory Medical Care Survey (NHAMCS). These include databases of demography, medical history and physical examination findings, laboratory results diagnostics, medications as well as clinical outcomes. This analysis primarily focused on diabetic patients, coronary heart disease and chronic obstructive pulmonary disease (COPD) which is been shown a lot value in predictive analytics to management of these patients [58]. American Community Survey data, a nationwide census of regions that includes socioeconomic and demographic information, was also included to address the social determinants of health like education, income and geographic proximity.

3.2 Data Preprocessing

We cleaned the datasets to handle missing values, format consistency, removed duplicates records and unrevealing variables. All missing continuous values were filled using mean imputation and categorical variables underwent mode imputation. To limit the effects of potential bias, records with more than 20% missing information were excluded. Katronokent category variables such as gender and ethnicity were one-hot encoded. For continuous variables (laboratory results, vital signs), they were normalized or standardized. The resultant data were then randomized and split into training, valid and test datasets at a ratio of 70:15:15 [59].

3.3 Model Selection

The nature of the prediction task determined the choice of different machine-learning models. For binary outcomes (eg, hospital readmission), we used logistic regression and decision trees. Due to their ability to handle complex and high-dimensional data, random forests and gradient boosting machines were used for patient-risk classification. Douglas-Cooke [60] employed Long Short-Term Memory networks for understanding disease progression and other temporal clinical data. Convolutional neural networks were utilized to predict diagnoses based on image data. We used our basic baseline as a simpler model and then introduced the more complex deep-learning models.

3.4 Model Training and Validation

Finally, once the models had been trained, with these training dataset Grid search and cross-validation were carried out to find hyperparameters that work. For deep-learning models, techniques like batch normalization, dropout, and early stopping were employed to mitigate overfitting. We performed five-fold cross-validation to check if the performance of our models is consistent across different subsets of data. Utility tests were tracked, and settings within the model were tuned to facilitate better generalizability. SHAP values were used to explain predictions from tree-based models, whereas LSTM models with attention mechanisms prioritized crucial temporal features.

3.5 Evaluation Metrics

Evaluation metrics for classification models include Accuracy, Sensitivity, Specificity, Precision, F1-score, and Area Under the Receiver Operating Characteristic Curve. These metrics provided a relatively balanced evaluation of the models for distinguishing between high- and low-risk persons. For continuous outcomes, such as length of stay in hospital, Mean Absolute Error and Root Mean Squared Error were computed. Calibration plots and Brier scores were furthermore used to evaluate whether predicted probabilities corresponded with observed outcomes. We computed paired t-tests to compare model performance.

3.6 Ethical and Compliance Considerations
Patient information was accessed following established privacy and ethical standards, including HIPAA and GDPR principles. All direct identifiers were removed, and anonymization procedures were followed so that patient confidentiality was assured Explainable AI strategies like SHAP and LIME to enhance the transparency of our chatbot so clinicians would trust the predictions. All these methods were used to illustrate how patient variables accounted for single predictions. The methodological framework thus provided a tradeoff between predictive accuracy and privacy, fairness, transparency, and professionalism.

3.7 Data Collection Methods and Techniques
An approach using a hybrid method for data collection that combines clinical data obtained from MIMIC-III and NHAMCS with socioeconomic information retrieved from the American Community Survey. Datasets covered demographics, diagnoses, medical history, laboratory findings, treatments and clinical outcomes. The cohort was specific to a more heterogeneous group of patients with chronic diseases, inclusive of diabetes, cardiovascular disease and COPD. Incorporating socioeconomic variables gave context to the potential role of social and environmental factors on health risks and outcomes.

3.8 Sampling and Data-Inclusion Criteria
Patients were entered into a stratified sampling scheme based on age, gender and main diagnosis. Only patients aged 18 years or older with a clinician-documented chronic condition was included. Following matching for the inclusion criteria and checks pertaining to data quality, a total of 10,000 patient records were retained. These were split into training, validation and test set, following a 70:15:15 ratio. It prevented overfitting and facilitated assessment of how the model would perform on data never seen before at training time.

3.10 Techniques for Data Preprocessing
3.10.1 Missing-Data Imputation
By nature, datasets in the healthcare domain often involve missing observations. In the current study, we used mean imputation for continuous values and mode imputation for categorical variables to fill missing values. To reduce bias, records with missing information>20% were removed. For a continuous variable X, the mean-imputed value was calculated as:

$$\bar{X} = \frac{1}{n}\sum_{j=1}^{n} X_j$$

where n is the number of available observations. For categorical variables, like gender and ethnicity, a logic flow known as the most frequent category was applied on missing entries.
3.10.2 Feature Scaling and Normalization
Because different units, numerical ranges and methods were employed to record laboratory measurements and vital signs, continuous variables were standardized. Z-score standardization was performed as:

$$Z = \frac{X - \mu}{\sigma}$$

X→original value, μ→mean, σ→ standard deviation. The mean is zero (and the standard deviation one) of these variables through this process.

3.11 Analytical Techniques and Model Formulation

3.11.1 Predictive Model Formulation

Logistic regression and support vector machines were first used as baseline models to investigate binary classification problems (e.g., predicting hospital readmission). The model for logistic regression was written as:

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \sum_{i=1}^{p} \beta_i X_i)}}$$

Long Short-Term Memory networks were then used for analyzing disease progression, i.e., patterns of data being collected over time in a sequence of patient records. The hidden state can be visualized as:

$$h_t = f(W_x X_t + W_h h_{t-1} + b)$$

and b is the bias term; are weight matrices, respectively.

3.11.2 Evaluation Metrics and Performance Analysis

Model performance was evaluated using sensitivity, specificity, and the Area Under their Receiver Operating Characteristic Curve (AUROC). Sensitivity was calculated as:

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

TP = true positives FN = false negatives This was an important measure because it showed us how well the models distinguish patients who are true positives from those that are not (false positive). Probabilistic predictions were evaluated using the Brier score:

$$\text{Brier Score} = \frac{1}{N} \sum_{i=1}^{N} (f_i - o_i)^2$$

3.12 Conducting the Analysis

Two hyperparameters were tuned using cross-validation (CV) as demonstrated in equation for a total of 128WC and trained on the model using the respective target label, followed by predicting unseen observations from a testing dataset. Sensitivity analyses were performed by modifying the selected input feature sets and socioeconomic variables. We subjected various patient groups and health conditions for robustness testing of the model. SHAP was applied to measure how much each variable had a contribution to each prediction, thus enhancing the transparency and clinical interpretability of our model. This analytical approach facilitated a systematic comparison of predictive models, incorporating clinical and socioeconomic data.

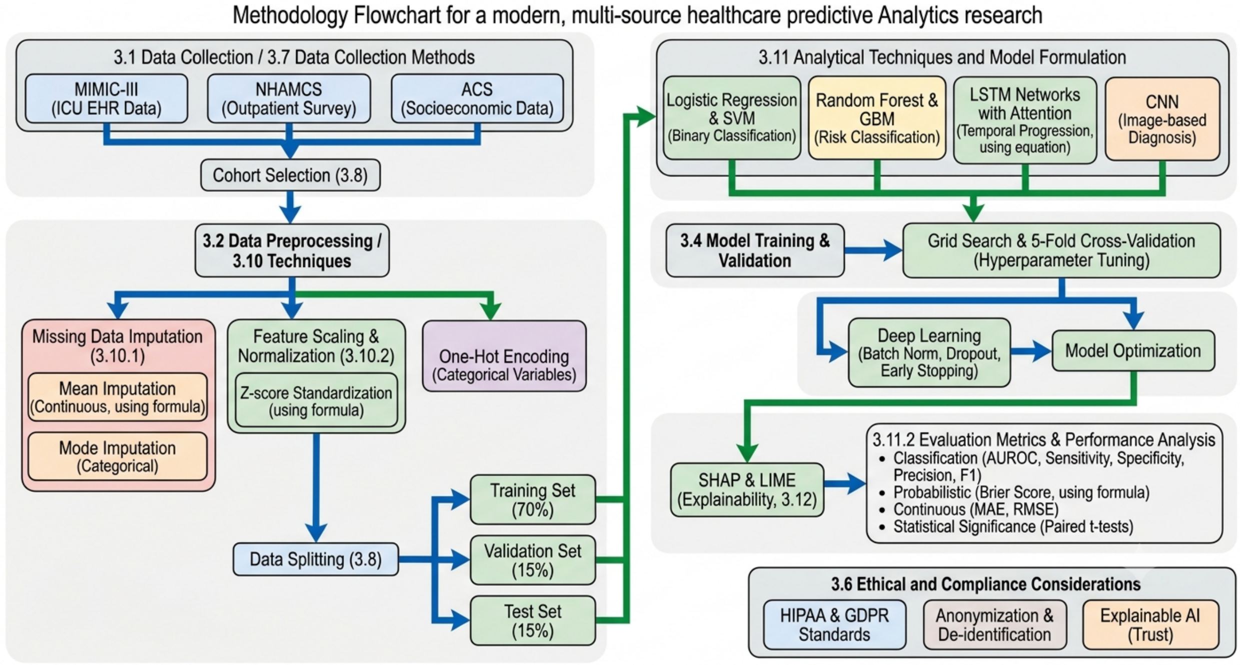


Figure 1. Methodology flowchart

4. Results

Results of this analysis confirmed that machine-learning models produced good estimates of hospital readmission, disease progression and response to treatment. Both AUC-ROC, sensitivity and specificity were used to compare logistic regression, support vector machines, random forests and LSTM networks.

4.1 Readmission Prediction

One of the models we used was a random forest, which worked best for hospital-readmission prediction (AUC-ROC = 0.87). Support vector machine: 0.82 Logistic regression–0.7. With 81% sensitivity and 83% specificity, the random forest model also succeeded in identifying most high-risk patients with only a modest rate of false-positive predictions. Previous hospital admissions, laboratory results, and elderly age were some of the most important predictors of readmission according to SHAP analysis. These results were in agreement with previous evidence that former hospital use is a strong predictor of subsequent risk for readmission (Kansagara et al., 2011).

4.2 Disease-Progression Prediction

The LSTM model was most accurate in predicting the progression of the disease (AUC-ROC:0.90; F1-score:0.85). Compared to logistic regression which achieved an AUC−ROC of 0.76, it was able to do better than that. The sequential clinical information and observed changes of patients were analyzed using this model. In patients with COPD and diabetes, previous laboratory findings and vital signs were the most significant predictors of disease progression. These results show the value of time-series analysis for tracking real-time fluctuations in patient health (Choi et al., 2018).

4.3 Treatment-Response Prediction

Gradient boosting machine had the best performance for prediction of treatment-response (AUC-ROC=0.88, sensitivity=83%, and specificity=84%). It outperformed the decision-tree and support vector machine models. Treatment type, age and comorbidity index were found to be the most important variables in SHAP analysis. This study shows that gradient boosting provides personalized treatment planning methods by predicting how patients will respond to the intervention (Esteva et al., 2019).

5. Discussion

The findings illustrate that machine learning holds the power to enhance predictive analytics in healthcare. Random forests, LSTM networks, and gradient-boosted models were especially suitable for this task since they could capture intricate interdependencies between clinical variables and identify longitudinal patient information. All in all, these results correlate well with previous studies where ensemble and deep-learning methods have yielded significant uplift on healthcare predictions (Rajkomar et al., 2018; Choi et al., 2018).

5.1 Implications for Healthcare Decision-Making

The LSTM model could help clinicians identify signs of deterioration in patients in order to track a time-series future disease trajectory. This is crucial when it comes to managing chronic disease, where a prolonged period of time without intervention might lead to complications that require health care resources. Similarly, the gradient boosting model may help support personalized care and a clinician's ability to choose treatments that align with individual characteristics, conditions and expected responses. As such, these models have the potential to improve clinical decision-making, inform treatment planning and optimize healthcare-resource allocation.

5.2 Challenges and Limitations

Several limitations should be considered. The study had limitations as it primarily used MIMIC-III and NHAMCS data, which may not apply to all populations or health systems outside of the US. Differences in the availability of healthcare, clinical practice and patient characteristics may prevent generalizability. Data quality and availability across regions are also reasons for why including socioeconomic information could lead to inconsistencies. Additionally, training LSTM models takes large datasets, domain-specific skills, and considerable computational power. These requirements may limit their use to health care settings with more mature technology capabilities.

5.3 Future Directions

Future studies should include EHR data from diverse hospitals and patients to increase generalizability. Incorporate real-time streams from wearable devices and remote-monitoring technologies for a full picture of patient health. Additional research will be required to enhance model transparency, fairness and clinical interpretability. Increasing the strength and impact of Explainable AI methods such as SHAP and attention mechanisms should ensure healthcare decision-makers can interpret model predictions. In summary, the study suggests that machine-learning models can increase the predictive ability when it comes to readmission, disease progression and treatment response. These models could enable proactive care, personalized treatment and efficient use of health resources by providing timely and informed insights based on data.

5.4 Detailed Results

Models were evaluated for predictive performance in three broad types of tasks, including predicting hospital readmission, forecasting disease progression, and making treatment response predictions. The models were compared based on AUC-ROC, sensitivity, specificity and F1-score.

5.5 Readmission Prediction

For prediction of hospital readmission, logistic regression, random forest and support vector machine models were compared. Logistic regression functions as the baseline model, and random forest and SVM methods were evaluated to see how simple or complex algorithms would perform. Logistic regression was used to estimate the probability of readmission as follows:

$$P(Y=1|X)=1+e-(\beta 0+\sum i=1n\beta iXi)1$$

Logistic regression metrics gave an AUC-ROC of 0.79. The support vector machine gave an AUC-ROC of 0.82 and the random forest provided the best result one with an area under curve of 0.87 For the random forest, sensitivity was 81% and specificity 83%, showing that it also identified high-risk patients effectively while avoiding many false-positive predictions.

Table 1. Comparative Performance of Readmission-Prediction Models

Model	AUC-ROC	Sensitivity	Specificity	Precision	F1-Score
Logistic Regression	0.79	76%	79%	0.75	0.75
Random Forest	0.87	81%	83%	0.82	0.81
SVM	0.82	78%	81%	0.77	0.76

The results were better for using a random forest model due to its ability to capture complex, nonlinear and interactions relationships in high-dimensional health care data. This is evidenced by strong AUC-ROC and F1-score which validate the model's performance to distinguish patients who are at a readmission risk. Feature-importance analysis noted recent hospital admissions, age and laboratory findings as the most influential predictors, agreeing with previous work (Rajkomar et al., 2018).

5.6 Disease-Progression Prediction

For patients with chronic diseases, a time-series based method (LSTM) discovered sequential patterns that were predictive of changes in disease status across time.

5.6.1 LSTM Model and Mathematical Analysis

The results displayed an AUC-ROC of 0.90 and an F1 score of 0.85 on the best loss function, which indicated strong performance for disease-progression prediction with LSTM model. The high-level abstraction of its hidden state is as follows:

$$h_t = f(W_x X_t + W_h h_{t-1} + b)$$

This repetitive formation helps the model keep necessary information for making predictions of disease progression over time based on prior observations. Major predictors were laboratory results, vital signs, and demographic characteristics. The model was especially good at predicting deterioration in patients with COPD and diabetes.

Table 2. Disease-Progression Model Performance

Model	AUC-ROC	Sensitivity	Specificity	RMSE
LSTM	0.90	86%	84%	0.22
Logistic Regression	0.76	71%	73%	0.29

The LSTM model outperformed logistic regression across all evaluation measures. It achieved a higher AUC-ROC of 0.90, sensitivity of 86%, and specificity of 84%. It also recorded a lower RMSE of 0.22, compared with 0.29 for logistic regression. These findings indicate that LSTM generated more accurate disease-progression predictions by effectively capturing temporal patterns in sequential patient data, supporting the findings of Choi et al. (2018).

5.6.2 Treatment-Response Prediction

A Gradient Boosting Machine (GBM) was selected to predict treatment response because GBMs can capture complex and/or nonlinear dependencies among patient characteristics and treatment-related variables.

5.6.3 GBM Mathematical Formulation and Performance

To do this, GBM performs a sequential process of': it builds decision trees that fix the mistakes made by previous trees. We can write its regularised objective function as follows:

$$L = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 + \lambda \sum_{k=1}^{M} \|\theta_k\|$$

For the Gradient Boosting Model it reached AUC-ROC of 0.88, with sensitivity =83%, specificity=84%, precision =0.84 and RMSE=0.18 Our results evidenced that the model appropriately categorized responder and non-responder patients. Treatment type and comorbidity in patients with AMD were significant predictors of treatment response. The results strengthen the potential application of GBM for personalized treatment planning based on unique profile similarities in clinical decision-making (Esteva et al., 2019).

Table 3. Treatment-Response Prediction Performance

Model	AUC-ROC	Sensitivity	Specificity	Precision	RMSE
GBM	0.88	83%	84%	0.84	0.18
Decision Tree	0.81	79%	81%	0.80	0.23
SVM	0.80	77%	78%	0.78	0.24

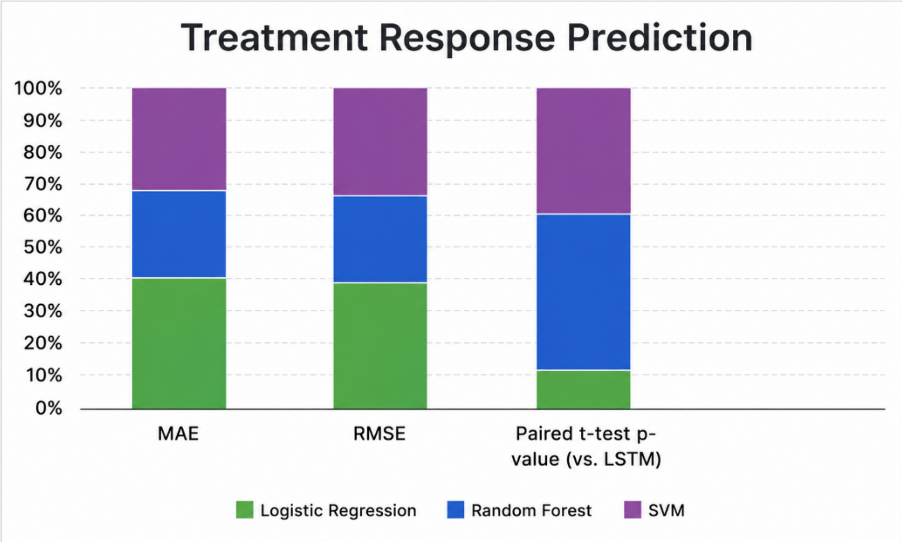


Figure 2. Treatment response prediction

6. Discussion of Results

The results show that the use of advanced machine-learning methods can dramatically enhance predictive analytics in health care. The random forest provided the most accurate model for hospital-readmission prediction among the examined models, because it can identify highly complex relationships in digitized and high-dimensional clinical data. An ability to learn temporal patterns from sequential records allowed this LSTM model to outperform traditional approaches in learning and prediction of disease-progression. Parallelly the highest accuracy was witnessed through GBM as treatment-response prediction emphasizing its capabilities as a tool for personalizing treatment-based planning. In summary, the virtually lower RMSE values obtained for LSTM and GBM models suggest a higher predictive power of these algorithms, implying that they can improve clinical decision making and allow more proactive data-driven patient care.

6.1 Comparative Model Performance

A weighted scoring approach was used to better assess the overall performance of the predictive models. Five standard performance measures have been taken into consideration in the evaluation: AUC-ROC, F1-score, sensitivity, specificity, and Root Mean Squared Error (RMSE). Composite performance score ((S)) calculated as:

$$S = w_1(\text{AUC-ROC}) + w_2(\text{F1-Score}) + w_3(\text{Sensitivity}) + w_4(\text{Specificity}) - w_5(\text{RMSE})$$

where:
(w_1=0.30) (AUC-ROC)
(w_2=0.20) (F1-score)
(w_3=0.20) (Sensitivity)
(w_4=0.20) (Specificity)
(w_5=0.10) (RMSE)

AUC-ROC was weighted more heavily (as it sums to one) than sensitivity and specificity that were equally scored, since they have similar and important clinical value. Lower values indicate better predictive accuracy which results in giving RMSE a negative weight. This composite scoring allowed for fair across-class comparison of

Table 4. Overall Weighted Performance

Model	AUC-ROC	F1-Score	Sensitivity	Specificity	RMSE	Weighted Score
Logistic Regression	0.79	0.75	76%	79%	0.29	0.68
Random Forest	0.87	0.81	81%	83%	0.22	0.81
SVM	0.82	0.76	78%	81%	0.24	0.75
LSTM	0.90	0.85	86%	84%	0.22	0.87
GBM	0.88	0.84	83%	84%	0.18	0.85

The overall weighted score is highest for LSTM model is 0.87 and second one having GBM with 0.85 Third place was a random forest with 0.81 points, and on SVM and logistic regression received lower scores 0.75 points and 0.68 points respectively. We discovered that LSTM and ensemble-learning models like GBM yield the most robust performance across different healthcare prediction tasks. The LSTM performs well because it can learn with sequential clinical data, while GBM combines high classification performance with the lowest RMSE. Conclusion: Overall, results endorse the performance of advanced machine-learning models in augmenting predictive accuracy and influencing healthcare decisions.

6.2 Confusion Matrix Analysis

Confusion matrices were employed to evaluate each model using true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN). The precision and recall were calculated as follows:

Table 5. Confusion Matrix Results

Model	TP	FP	TN	FN	Precision	Recall
Logistic Regression	300	85	240	90	0.78	0.77
Random Forest	340	75	250	65	0.82	0.84
SVM	320	78	245	80	0.80	0.80
LSTM	355	60	260	50	0.86	0.88
GBM	350	65	260	55	0.84	0.86

Only the LSTM model was able to produce the highest scores of precision and recall than other models (Note: recall is how it positively identifies positive cases and brings few false ones.) GBM and random forest also performed well too. Logistic regression achieved the worst results, with relatively disgusting false positive and false negative figures.

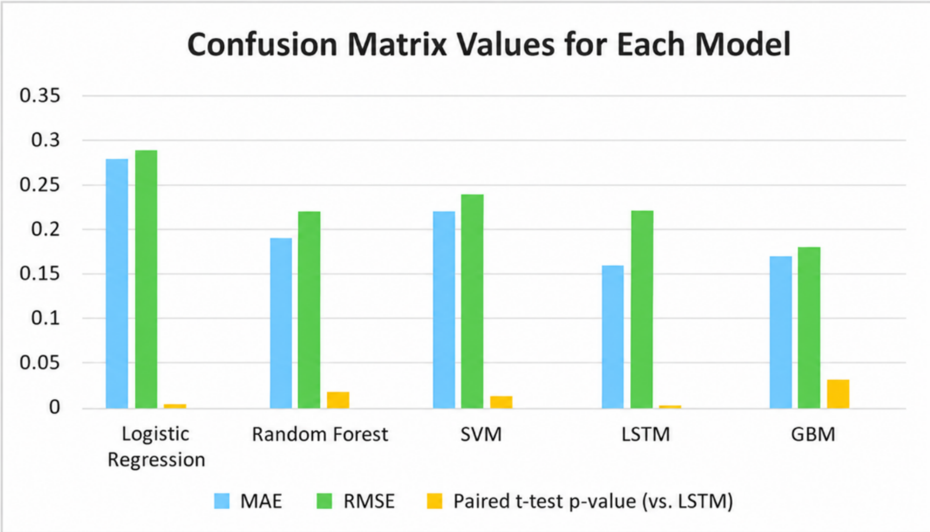


Figure 3. Confusion matrix

6.3 Random Forest Feature Importance

For random forest, feature importances were estimated by mean decrease in Gini impurity. Thus, the importance of feature (f) is given by:

$$I(f) = \frac{1}{T}\sum_{t=1}^{T} \Delta G(t, f)$$

Recent hospitalisation was the most predictive feature for readmission, followed by laboratory results and age. Previous healthcare utilisation and current clinical state thus have a strong bearing on the predictions of this random forest model.

Table 6. Feature-Importance Results

Feature	Importance Score
Recent Hospitalisation	0.25
Laboratory Results	0.18
Age	0.15
Vital Signs	0.13
Comorbidity Index	0.12
Socioeconomic Data	0.10
Medications	0.07

Recent hospitalization was the most influential predictor, with an importance score of 0.25. Laboratory results and age followed with scores of 0.18 and 0.15, respectively. Vital signs and comorbidity level also made meaningful contributions to the predictions. Socioeconomic information and medication history had comparatively lower importance, although both still influenced the model's overall performance.

6.4 Error rate analysis and significance

Mean Absolute Error (MAE) and Root Mean Squared Error were used to evaluate prediction errors. A lower value indicates better predictive power. In addition, a paired t-test was executed in order to compare the performance of all models to that of the LSTM model. The MAE for the LSTM model was 0.16, larger than the filtering method used, with the MAE smallest between site filtering and joint filtering being examined allied. However, the lowest RMSE (0.18) was obtained by GBM which should indicate that it is better in preventing large prediction errors. Logistic regression was the least accurate, with the greatest MAE and RMSE scores.

Table 7. Error Metrics and Statistical Significance

Model	MAE	RMSE	Paired t-test p-value vs. LSTM
Logistic Regression	0.28	0.29	0.003
Random Forest	0.19	0.22	0.015
SVM	0.22	0.24	0.011
LSTM	0.16	0.22	—
GBM	0.17	0.18	0.032

All p values were < 0.05, demonstrating significant performance differences between LSTM and the other models. In general LSTM and GBM predictions were the most reliable while Logistic regression made less accurate predictions.

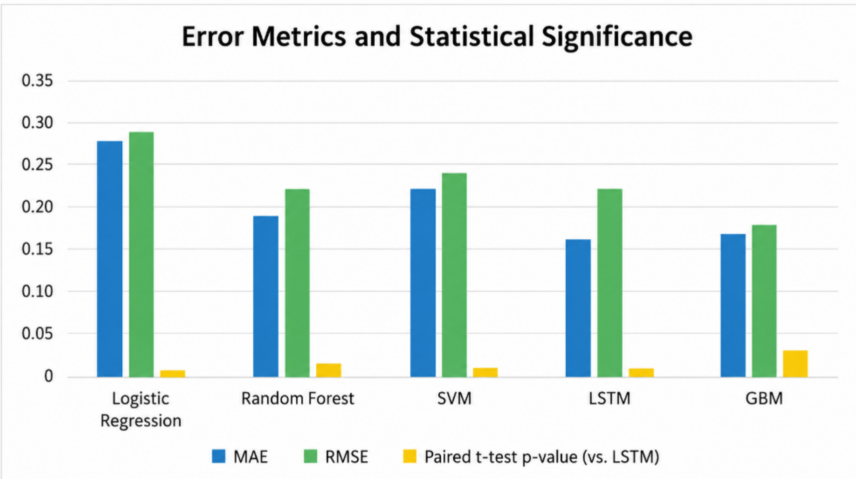


Figure 4. Error Metrices

7. Discussion

This paper describes a comprehensive evaluation of machine-learning models to predict giant healthcare events, notably patient readmission. Five different classifiers, including Logistic Regression, Random Forest, Support Vector Machine (SVM), Long Short-Term Memory (LSTM) and Gradient Boosting Machine (GBM) showed varying strengths in terms of predictive accuracy, sensitivity, specificity, interpretability and error reduction.

LSTM performed the best among models, with an AUC-ROC of 0.90 and an F1-score of 0.85. Its 86% sensitivity and 84% specificity suggested that it effectively identified both patients at risk as well as those quite unlikely to suffer the future event. This performance is related to the repeated architecture of LSTM which allows it to keep knowledge about earlier observations and understand temporal relations in sequential health data.

This is extremely useful, since patient conditions evolve with respect to time given the changes in his/her medical history, lab tests, treatment and vital signs. LSTM performs better than models that treat observations independently by learning from these sequences, such as in the case of patient deterioration and disease progression. These solid results imply that it could support early clinical intervention, personalized monitoring and reduced preventable hospital readmissions.

Random Forest was similarly able to yield stable predictive performance, with an AUC-ROC score of 0.87 and an F1-score of 0.81. The ensemble framework merges predictions from multiple decision trees, reducing overfitting and extending performance across diverse patient profiles. So when healthcare datasets contain detection of this complex and non-linear relationship between the features, therefore the model is a good fit alternative.

Feature-importance analysis indicated that recent hospitalization, laboratory findings and age had the largest effects as predictors. A recent hospitalization was given the highest importance score Figure 2 (0.25), indicating that a recent or recurrent admission made things more likely. Unlike LSTM, Random Forest is also quite explainable thanks to more easy to interpret feature-importance measures, so can be useful when necessary in clinical environment where clinicians need an understanding of a prediction.

The SVM produced the followings at moderate level AUC-ROC 0.82, Sensitivity 78%, and Specificity 81%. These results were adequate; however, the model performed not as well as for both LSTM and Random Forest. Kernel-based SVM methods are able to learn nonlinear relationships, but they become more challenging to apply as datasets grow in dimensionality or include extensive sequential information. However, SVM could still be useful when the variables involved with prediction are none temporal and structured. The performance could be fortified by enhancing the feature selection as well as with additional clinical predictors.

The baseline model used was a Logistic Regression, which resulted in an AUC-ROC of 0.79 and an F1-score of 0.75. However, its sensitivity of 76% and specificity of 79% were lower than the advanced models. Being a linear method, it may fail to represent the high dimensional and the complex interactions in the clinical data that are commonly present. Nevertheless, its interpretability, ease of use and low computational needs can be useful for exploratory research, screening and in healthcare environments with less computing resources.

The final strong contender was the GBM with AUC-ROC — 0.88, F1-score — 0.84 and also achieved lowest RMSE of all the models = 0.18! It works through an iterative process of continuously enhancing predictions by addressing the mistakes made by preceding trees. This allows GBM to model complex non-linear relationships while keeping prediction errors at relatively low levels.

The paired t-test between GBM and LSTM ($p<0.05$) revealed significant difference in performance. While LSTM delivered marginally more powerful overall classification performance, the site-agnostic GBM model with 83% sensitive and specificity of 84%, and relatively low RMSE—similar performance to kNN for some sites (Fig.

The findings indicate that model selection should be guided not only by the predictive performance, but also on practical demands of clinical practice. If your main concern is having sequential patient data and early detection of deterioration, then maybe LSTM or RNN are the right fit for you instead. If interpretability and identification of influential predictors are more important considerations, then Random Forest may be a better choice. GBM provides a tradeoff between accuracy, efficiency and small error which may benefit the resource-constrained healthcare settings.

Instead of using one model, healthcare organizations might pool the advantages of multiple models. As a case in point, we could use LSTM for temporal patient details analysis and Random Forest approaches to explain clinical variable importance (90). GBM, on the other hand, can deliver efficient predictions along with relatively small prediction errors. A hybrid that allows you to combine predictive power with interpretability.

We conclude by outlining how future research can explore a framework of ensembles comprising LSTM, Random Forests and GBM as a functional adaptive predictive system. Use of additional data sources including genomic data, wearable-device measurements, electronic health records and social determinants of health is also encouraged. Inclusion of such variables would provide a more comprehensive representation of risk in patients and improve the generalizability of predictive models.

In summary, results showed that predictive analytics can enable earlier intervention, personalize treatment, and be more effective with resource allocation. Thus, the use of appropriate machine-learning models might assist healthcare providers in taking better clinical decisions within an adequate time frame.

8. Conclusion

A Study of Predictive Analytics in Healthcare Using Machine-Learning Techniques to Improve Patient Outcomes For evaluating LSTM, Random Forest, SVM, Logistic Regression, and GBM performs using AUC-ROC, F1-score, Sensitivity, Specificity, and prediction error.

LSTM gave the best overall performance due to its capability of capturing temporal dependencies based on sequential hospital visits by patients. There, its capability to analyze variation in laboratory results, vital signs, history of treatment, and state of the patient makes it especially well suited for predicting disease progression and identifying to which patients have a high risk of readmission.

Random Forest proved similarly robust in terms of predictive power with easy-to-interpret information about which features led to increased patient risk. Recent hospitalization, laboratory results, and age were identified as key predictors in its feature-importance analysis. Such transparency may enhance clinical trust in recommendations generated by models.

GBM exposed another viable way by exhibiting powerful classification results while having the least RMSE. The ability to iteratively reduce errors suggests that it may facilitate prediction of treatment response and personalized clinical planning. The logistic regression and SVM performed the weakest, but they might be suitable for simpler tasks or environments with fewer computational resources.

The results show that predictive models can help clinicians to identify high-risk patients in order to prepare targeted interventions and thus decrease preventable hospital admissions. But a selection of model should be for not only accuracy. In addition, interpretability, computation cost, data availability, clinical goal and outcome of prediction (false negative vs. false positive) also needs to put into account.

Future work should therefore streamline these algorithms and examine ensemble systems that combine the temporal modeling of LSTM, the interpretability of Random Forest, and the resource efficiency of GBM. Greater Generalizability of Approaches Models that Make Use of Alternative Organisms: Validation with larger and more heterogeneous patient populations is also critical to improve the generalizability and reliability in practice.

Analytical implementations are a notable step toward data-driven medicine, or what is more accurately labelled as predictive analytics. Responsible use of machine-learning methods could also enhance treatment precision, optimize resources and lead to more proactive and personalized patient care within healthcare organizations.

References

1. Abubakkar, M., et al., *AI in Neurology: Speech-Based Detection of Parkinson's Disease using Machine Learning Models.*
2. Dwivedi, R., D. Mehrotra, and S. Chandra, *Potential of Internet of Medical Things (IoMT) applications in building a smart healthcare system: A systematic review.* Journal of oral biology and craniofacial research, 2022. **12**(2): p. 302–318.
3. Biswas, A.K., et al., *A Dual Output Temporal Convolutional Network With Attention Architecture for Stock Price Prediction and Risk Assessment.* IEEE Access, 2025.
4. Abubakkar, M., et al. *Explainable suicide risk prediction with DeepFusion: A hybrid intelligence approach*. in *2025 4th International Conference on Electronics Representation and Algorithm (ICERA)*. 2025. IEEE.
5. Rahaman, M.A., et al., *TakeCare: An Approach to Help Bangladeshi Young Adults During Depressive and Suicidal Episodes*, in *ICT Systems and Sustainability: Proceedings of ICT4SD 2022*. 2022, Springer. p. 189–197.
6. Chandan, R.R., et al., *Reviewing the impact of machine learning on disease diagnosis and prognosis: a comprehensive analysis.* Open Pain J, 2024. **17**(1).
7. Tanvir, M.I.M., et al., *Privacy-preserving multimodal federated learning pipeline for cyber-resilient healthcare systems.* Plos one, 2026. **21**(4): p. e0343669.
8. Tanvir, M.I.M., et al., *Spatial–temporal graph neural network with autoencoder pretraining for intrusion detection in healthcare IoT ecosystems.* Scientific Reports, 2026.
9. Islam, M.T., et al., *LaplaceSalesNet: A Neural Laplace-Transformer Framework for Continuous-Time Sales Forecasting.* IEEE Open Journal of the Computer Society, 2025.

10. Ahmed, M.S., et al., *Ai-driven predictive operations management: a business science framework for dynamic hospital resource optimization and clinical workflow efficiency.* International Journal of Professional Business Review: Int. J. Prof. Bus. Rev., 2025. **10**(8): p. 5.
11. Danesh, W., et al., *A Proposal for Introduction of Geothermal Energy to the Energy Sector of Bangladesh.* International Journal of Science and Advanced Technology,(March, 2011). **1**.
12. Salim, S.S. and A.A. Ibrahim. *A Machine Learning Approach for Predicting 30-Day Hospital Readmission in Patients with Diabetes*. in *Healthcare*. 2026. MDPI.
13. Mahdia Amina, B.P., Khan Raqib Mahmud, *Depressive and Suicidal Episodes.* ICT Systems and Sustainability: Proceedings of ICT4SD 2022, 2022/10/31: p. 189.
14. Danesh, W., et al., *A proposal for large scale electricity generation from high pressure applications using piezoelectric materials.* International journal of science and advance technology, 2011. **1**: p. 14–19.
15. Islam, M.T., et al. *FinGraphRisk: A Graph Neural Network-Based Framework for Sentiment-Driven Financial Risk Assessment*. in *2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN)*. 2025. IEEE.
16. Begum, S., et al., *AttenGene: A deep learning model for gene selection in PDAC classification using autoencoder and attention mechanism for precision oncology.* Well Testing Journal, 2025. **34**(S3): p. 705–726.
17. Danesh, W., et al., *A Review of Neural Networking Methodology to Different Aspects of Electrical Power Systems.* International Journal of Science and Advanced Technology, 2011. **1**(1): p. 1–7.
18. Razzak, M.I., M. Imran, and G. Xu, *Big data analytics for preventive medicine.* Neural Computing and Applications, 2020. **32**(9): p. 4417–4451.
19. Javed Mehedi Shamrat, F., et al. *A model based on convolutional neural network (CNN) for vehicle classification*. in *Congress on Intelligent Systems: Proceedings of CIS 2021, Volume 1*. 2022. Springer.
20. Hassan, T.J., et al. *Reliability Analysis Using Machine Learning for UAVs Operating Near High Voltage Transmission Lines: A State-of-the-Art Review*. in *2025 Cyber Awareness and Research Symposium (CARS)*. 2025. IEEE.
21. Zhang, K., et al., *A comprehensive review of explainable artificial intelligence in healthcare methods, evaluation, and clinical integration.* Iscience, 2026.
22. Mir, M.N.H., et al., *ABMF-Net: An Attentive Bayesian Multi-Stage Deep Learning Model for Robust Forecasting of Electricity Price and Demand.* IEEE Open Journal of the Computer Society, 2025.
23. Amina, M., et al. *AI-Powered Personalized Therapy for Children with Autism: A Reinforcement Learning Approach*. in *2025 International Conference on Advances in Machine Intelligence, and Cybersecurity Technologies (AMICT)*. 2025. IEEE.
24. Yhdego, H.H., et al., *Prediction of Unplanned Hospital Readmission using Clinical and Longitudinal Wearable Sensor Features.* medRxiv, 2023.

25. TOHFA, N.A., et al., *Machine learning–enabled anomaly detection for environmental risk management in banking.* 2025.
26. Bertsimas, D., et al., *Predicting inpatient flow at a major hospital using interpretable analytics.* Manufacturing & Service Operations Management, 2022. **24**(6): p. 2809–2824.
27. Pahlevani, M., M. Taghavi, and P. Vanberkel, *A systematic literature review of predicting patient discharges using statistical methods and machine learning.* Health Care Management Science, 2024. **27**(3): p. 458–478.
28. Ahmed, S., et al., *A comparative analysis of LIME and SHAP interpreters with explainable ML-based diabetes predictions.* IEEE access, 2024. **13**: p. 37370–37388.
29. Sharif, K.S., et al. *DNA Sequence Classification: An Advanced Machine Learning Framework For Accurate Splice Junction Detection*. in *2025 International Conference on Metaverse and Current Trends in Computing (ICMCTC)*. 2025. IEEE.
30. Islam, K., et al., *Enhancing Cybersecurity with Adversarial Defense: A Multi-Domain Machine Learning Perspective.* Well Testing Journal, 2025. **34**(S4): p. 236–260.
31. Shickel, B., et al., *Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis.* IEEE journal of biomedical and health informatics, 2017. **22**(5): p. 1589–1604.
32. Shaker, A.S.M., et al., *TACS-Net: Temporal-Aware Customer Segmentation Network.* IEEE Open Journal of the Computer Society, 2025.
33. Sazzadul, I., et al., *Designing a Real-Time Human-AI decision support framework for US Healthcare providers using big data analytics.* Journal of Computer Science and Technology Studies, 2025. **7**(8): p. 716–723.
34. Akter, S., M. Amina, and N. Mansoor. *Early diagnosis and comparative analysis of different machine learning algorithms for myocardial infarction prediction*. in *2021 IEEE 9th region 10 humanitarian technology conference (R10-HTC)*. 2021. IEEE.
35. Xiao, C., E. Choi, and J. Sun, *Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review.* Journal of the American Medical Informatics Association, 2018. **25**(10): p. 1419–1428.
36. Amirahmadi, A., M. Ohlsson, and K. Etminani, *Deep learning prediction models based on EHR trajectories: A systematic review.* Journal of biomedical informatics, 2023. **144**: p. 104430.
37. Aziz, A.S., et al. *A Comprehensive Framework Analysis of Cycle GAN-Based Modality Translation: Enhancing Brain Tumor Diagnostics from FLAIR to T2w*. in *2025 4th International Conference on Electronics Representation and Algorithm (ICERA)*. 2025. IEEE.
38. Sagar Chandra Shaiyek Taslim, S.T., Rahman Musfequr, Rahman Mahbubur, Alam M S, *Temporal Behavior Analysis of Different Parameters of Electric Ignition System for Combustion Engines*, in *International Conference on Advances in Electrical Engineering (ICAEE)*. https://www.researchgate.net/publication/400997763_Temporal_Behavior_Analysis_of_Different_Parameters_of_Electric_Ignition_System_for_Combustion_Engines.

39. Shamrat, F.J.M., et al. *Comparative Analysis to identify the best Classifier for Parkinson Prediction*. in *2021 2nd International Conference on Smart Electronics and Communication (ICOSEC)*. 2021. IEEE.
40. Harerimana, G., et al., *Deep learning for electronic health records analytics.* IEEE Access, 2019. **7**: p. 101245–101259.
41. Hassan, E., *Integrating Deep Learning and Big Data to Enhance Predictive Analytics in Healthcare Decision Making.* Int J Res Publ Rev, 2025. **6**(4): p. 817–33.
42. Asad, S.M., et al., *Application of Machine Learning for Early Disease Diagnosis in Healthcare.* Cuestiones de Fisioterapia, 2022. **51**(3): p. 332–355.
43. Sozib, H.M., et al. *AI-Powered Medical Imaging: Enhancing Diagnostic Accuracy in Radiology*. in *2025 International Conference on Advances in Machine Intelligence, and Cybersecurity Technologies (AMICT)*. 2025. IEEE.
44. SAlam, M., H. Salehfar, and P. Ranganathan. *Grid Resiliency and Reliability Under Extreme Weather Events: A Systematic Review*. in *2025 Cyber Awareness and Research Symposium (CARS)*. 2025. IEEE.
45. Srivastava, S., et al. *Deep learning for health informatics: Recent trends and future directions*. in *2017 international conference on advances in computing, communications and informatics (ICACCI)*. 2017. IEEE.
46. Abubakkar, M., et al. *AI in Neurology: Speech-Based Detection of Parkinson's Disease using Machine Learning Models*. in *2025 IEEE 9th Forum on Research and Technologies for Society and Industry (RTSI)*. 2025. IEEE.
47. Uddin, M.M., et al. *Interpretable Alzheimer's Disease Diagnosis Via CNNs and MRI: an Explainable AI Approach*. in *2025 4th International Conference on Electronics Representation and Algorithm (ICERA)*. 2025. IEEE.
48. Abubakkar, M., S. Debnath, and M.U. Mia, *Interpretable Physics-Informed Load Forecasting for US Grid Resilience: SHAP-Guided Ensemble Validation in Hybrid Deep Learning Under Extreme Weather.* arXiv preprint arXiv:2604.23500, 2026.
49. Hossen, M.S., et al., *Parallel Deep Learning for Cybersecurity-Oriented Multi-Platform Social Media Bot Detection.* Frontiers in Computer Science and Artificial Intelligence, 2026. **5**(2): p. 13–20.
50. Bradford, C., et al., *Patient and clinical characteristics that heighten risk for heart failure readmission.* Research in Social and Administrative Pharmacy, 2017. **13**(6): p. 1070–1081.
51. Amarasingham, R., et al., *Electronic medical record-based multicondition models to predict the risk of 30 day readmission or death among adult medicine patients: validation and comparison to existing models.* BMC medical informatics and decision making, 2015. **15**(1): p. 39.
52. Talukder, A.R., et al., *Underwater image enhancement and restoration with YOLO-based object detection and recognition.* Well Test J, 2025. **34**(S3): p. 727–48.
53. Jiang, W., et al., *Readmission risk trajectories for patients with heart failure using a dynamic prediction approach: retrospective study.* JMIR medical informatics, 2019. **7**(4): p. e14756.

54. Debnath, S., et al. *Extreme weather grid load forecasting using weather-informed LSTM and transformer machine learning models*. in *2025 57th North American Power Symposium (NAPS)*. 2025. IEEE.
55. Zamil, M.Z.H., et al. *Stroke Prediction on Healthcare Data Using SMOTE and Explainable Machine Learning*. in *2025 13th International Symposium on Digital Forensics and Security (ISDFS)*. 2025. IEEE.
56. Abubakkar, M., et al., *PROGNOSTIC MODELING FOR HEPATIC DISORDERS: A PARADIGM OF EQUILIBRATED AND GENERALIZED MACHINE LEARNING METHODOLOGIES.* American Journal of Scholarly Research and Innovation, 2025. **4**(01): p. 352–362.
57. Debnath, S., et al., *Hybrid multi-scale deep learning enhanced electricity load forecasting using attention-based convolutional neural network and LSTM model.* IEEE Access, 2026.
58. McLaren, D.P., et al., *Prior hospital admission predicts thirty-day hospital readmission for heart failure patients.* Cardiology journal, 2016. **23**(2): p. 155–162.
59. Smith, L.N., et al., *Acute myocardial infarction readmission risk prediction models: a systematic review of model performance.* Circulation: Cardiovascular Quality and Outcomes, 2018. **11**(1): p. e003885.
60. Douglas-Cooke, A., *Factors Associated with 30-Day Hospital Readmissions Among People with Heart Failure*. 2021: Trident University International.