Advances in Nonlinear Variational Inequalities
ISSN: 1092-910X
Vol 27 No. 3 (2024)

Advancements in Machine Learning Algorithms for Predictive Analytics in Healthcare

Dr. Pranoti Prashant Mane[1], Dr. Ketaki Naik[2], Lalit R. Chaudhari[3] Dr. H. E. Khodke[4], Dr. Vilas S. Gaikwad[5]

[1]Associate Professor and HOD, Department, Electronics & Telecommunications, MES's Wadia College of Engineering, Pune. ppranotimane@gmail.com

[2]Associate Professor, Department: of Information Technology,Bharati Vidyapeeth's College of Engineering for Women, Pune. ketaki.naik@bharatividyapeeth.edu

[3]Assistant Professor, Department Instrumentation Engineering , Dr. D. Y. Patil Institute of Technology ,Pune lali24006@gmail.com "

[4]Assistant Professor, Computer Engineering department, Sanjivani College of Engineering Kopargaon (An Autonomous institute), Maharashtra, India. khodkehecomp@sanjivani.org.in

[5]Associate Professor and HOD, Department of Information TechnologyTrinity College of Engineering and Research Pune. vilasgaikwad11@gmail.com

Article History:

Received: *20-02-2024*

Revised: *29-04-2024*

Accepted: *23-05-2024*

Abstract

Recent improvements in machine learning (ML) algorithms have changed predictive analytics in healthcare in a way that has never been seen before. This means that there are now more ways than ever to improve patient results and business efficiency. This summary looks at important changes and what they mean. A lot of healthcare data, like electronic health records (EHRs), medical images, genetics, and personal sensor data, is being used more and more with machine learning methods like guided learning, unsupervised learning, and deep learning. These methods make it possible to use predictive models to diagnose diseases, give patients specific treatment suggestions, and keep track of their care. For sorting things into groups, supervised learning techniques like support vector machines (SVM) and random forests have been used to correctly spot diseases based on complicated data trends. For example, SVMs have been useful for telling the difference between different types of cancer from genetic data, which helps with focused treatments. On the other hand, unsupervised learning algorithms like grouping algorithms help find groups of patients who share similar traits, which makes personalized medicine possible. Deep learning, has been very successful in medical picture analysis, being more accurate than humans at tasks like finding tumors in x-rays and lab slides. Its ability to instantly learn traits from raw data has made diagnosis easier and more accurate. ML algorithms also help healthcare operations run more smoothly by using prediction analytics to help hospitals handle their resources better, make the best use of their staff, predict which patients will need to be admitted, and lower the number of times they have to be readmitted. These predictive models use a variety of data sources to guess how patients will do and how they will use healthcare resources, which helps people make smart decisions and make the best use of their resources.

Keywords: Predictive Analytics, Machine Learning Algorithms, Healthcare, Personalized Medicine, Medical Imaging, Operational Efficiency

1. Introduction

In recent years, machine learning (ML) algorithms have been added to healthcare systems. This has started a new era of predictive analytics that could completely change how patients are cared for, how efficiently operations are run, and how diseases are managed. This introduction talks about the progress and effects of machine learning (ML) in healthcare predictive analytics. It focuses on important methods, problems, and possible uses. Data-driven methods are being used more and more in healthcare to improve results and make the best use of resources. It's not always easy to deal with the huge amount and complexity of healthcare data, such as that found in electronic health records (EHRs), medical images, genetic data, and real-time sensor data from smart tech [1], [2]. With its ability to look at huge amounts of data and find useful trends, machine learning has the potential to change the way healthcare is provided by allowing early disease diagnosis, individual treatment plans, and proactive health management. The methods for machine learning have changed a lot to deal with the unique problems that healthcare data presents [3]. A variety of supervised learning algorithms, including support vector machines (SVM), decision trees, and ensemble methods, have been used to classify diseases, predict risks, and predict how well treatments will work. These algorithms make guesses based on trends they find in named data that they learn from.

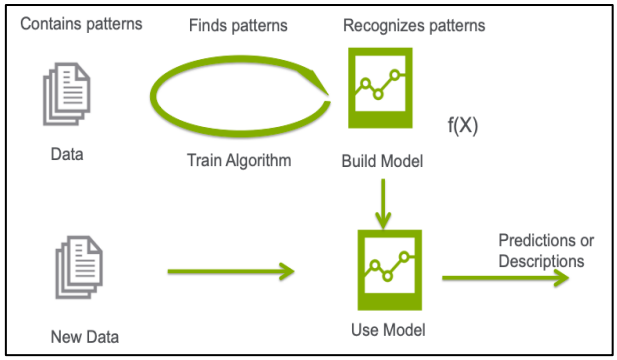

Figure 1: Overview of Model for Predictive analytics

Unsupervised learning methods like grouping and anomaly detection help divide patient groups into groups with similar health profiles and find outliers that could be signs of rare diseases or treatment not working [4]. These methods, illustrate in figure 1, are important for finding hidden trends in data without naming it first, which improves the accuracy of diagnosis and the effectiveness of treatment. Recent progress in deep learning, a type of machine learning that uses neural networks with many layers, has been very successful at tasks such as analyzing medical images, natural language processing (NLP) for clinical notes, and genetic sequence analysis. In [5] medical pictures, convolutional neural networks (CNNs) are great at pulling out features, while recurrent neural networks (RNNs) are great at handling sequential data like patient records or physiological data collected over time. ML has a lot of potential, but it's not easy to use in healthcare. Since health information is so private, data safety and security are very important. Ensuring a strong model's interpretability and explainability is important for clinical acceptance because medical workers need to be able to see and understand the choices that are made in healthcare. There are still big problems like not having enough high-quality tagged data and having to harmonize data from different sources [6]. It is common for healthcare datasets to have different data forms, quality, and completeness,

which means they need strong preparation methods and feature engineering that is specific to the topic. Also, the moral effects of computer bias and fairness in healthcare decision-making need to be looked at very closely. If there are biases in the training data or algorithms, there could be differences in how well different groups of patients do, which would make healthcare gaps worse instead of better [7].

2. Related Work

In healthcare, predictive analytics has become an important tool for better patient results, making the best use of resources, and making operations run more smoothly. This literature review looks at some of the most important studies and results in the field, with a focus on methods, uses, and results. In healthcare situations, predictive analytics has been used in a number of different ways. Using structured data from electronic health records (EHRs), supervised learning algorithms like logistic regression and support vector machines (SVM) have been used a lot to guess what will happen in clinical trials [8]. These programs look at a person's medical background, data, and biomarkers to guess how likely they are to get diseases like cancer, diabetes, and heart disease. Unsupervised learning methods, like grouping and anomaly identification, are used to divide patient groups into groups and find outliers in healthcare data. These methods are especially useful in personalized medicine, where they help make treatment plans that are based on the unique traits and health histories of each patient. Recent progress in deep learning, which uses neural networks with many layers, has changed the way medical picture analysis and natural language processing (NLP) jobs are done. Convolutional neural networks (CNNs) are very good at finding problems in medical pictures like X-rays and MRIs [9]. Recurrent neural networks (RNNs), on the other hand, use sequential data to guess how patients will do and how their diseases will get worse. Predictive analytics is used in many areas of healthcare, from diagnosing and predicting diseases to planning treatments and managing patients. One important use is predicting hospital readmissions. Predictive [10] models help healthcare workers step in early to stop needless readmissions, which improves patient care and lowers healthcare costs. In managing chronic diseases, predictive analytics helps find people who are at a high risk and could benefit from proactive measures like changing their lifestyle or medications. By guessing when a disease will get worse or cause problems, healthcare professionals can provide better care and use their resources more wisely.

Predictive analytics are used in population health management to find patterns and trends in big datasets. This lets public health officials and lawmakers put in place focused actions and preventative measures. The goal [11] of these activities is to improve the health of the community and lower differences in health care. Studies that looked at the effects of predictive analytics in healthcare found that it improved patient outcomes, lowered healthcare costs, and helped doctors make better decisions. Using predictive insights can help healthcare organizations get patients more involved and make them happier. It can also help them be more efficient and make better use of their resources. But there are still big problems, like worries about data protection, computer bias, and how to use prediction models in healthcare processes. Making sure that predictive analytics is used in an ethical way and that decision-making processes are open and clear are important for building trust between healthcare workers and patients.

Table 1: Summary of related work

Method	Algorithm	Key Finding	Limitation	Scope
Supervised Learning [12]	Support Vector Machines (SVM)	SVMs achieve high accuracy in diagnosing diseases from medical images.	Requires large labeled datasets for training.	Application in disease diagnosis and medical imaging.
	Random Forests	Effective in predicting patient outcomes based on diverse clinical data.	Prone to overfitting with noisy data.	Outcome prediction and treatment planning.
Unsupervised Learning [13]	K-means Clustering	Segmentation of patient cohorts with similar health profiles for personalized medicine.	Sensitivity to initialization parameters.	Patient clustering and cohort identification.
	Anomaly Detection	Detects rare diseases or adverse drug reactions from EHRs and sensor data.	Challenges in defining normal versus anomalous behavior.	Early detection of anomalies in patient health data.
Deep Learning [14]	Convolutional Neural Networks (CNNs)	Automates detection and classification of abnormalities in medical images.	Requires large computational resources and annotated datasets.	Medical imaging analysis and pathology detection.
	Recurrent Neural Networks (RNNs)	Predicts disease progression from sequential patient data in EHRs.	Difficulty in capturing long-term dependencies in data.	Longitudinal analysis and treatment response prediction.
Hybrid Approaches [15], [16]	Ensemble Methods	Combines multiple models for enhanced predictive accuracy and robustness.	Complex integration and interpretation of ensemble outputs.	Integration across diverse healthcare datasets for comprehensive analysis.
Natural Language Processing	Word Embeddings (e.g., Word2Vec)	Extracts semantic relationships from clinical notes for predictive modeling.	Limited interpretability of learned embeddings in medical context.	NLP-based clinical decision support systems.

3. Methodology

This proposed methodology outlines a systematic approach to applying machine learning (ML) algorithms for predictive analytics in healthcare, aiming to enhance patient outcomes, optimize resource allocation, and improve operational efficiency.

1. Data Acquisition and Preprocessing:

The part of collecting and preparing data is essential for building strong prediction models in healthcare. At first, different types of data are put together, from electronic health records (EHRs) to medical image files to smart devices to patient reports of results. Putting these sources together makes sure that the data is consistent and works with other data, which is important for a full study [17]. The next steps in cleaning data are to deal with missing numbers, errors, and standardizing forms to make the data better. Feature engineering is very important because it pulls out relevant traits that are clinically relevant and have predictive power. This careful process not only gets the

data ready for accurate modeling, but it also makes it possible to find insights that can be used right away to make big changes in healthcare decisions and patient results.

$$D_{integrated} = sum_{\{i=1\}}^{\{N\}} D\_i$$

This equation represents the integration of data from multiple sources (D_i) into a single comprehensive dataset (D_integrated), essential for aggregating diverse healthcare data types such as electronic health records (EHRs), medical imaging, and patient-reported outcomes.

$$D\_cleaned = D\_raw - NaNs(D\_raw)$$

D_cleaned is derived from D_raw by removing missing values (NaNs), ensuring data quality and completeness for subsequent analysis. Handling missing data is critical to prevent bias and inaccuracies in predictive models.

$$D\_outliers = Clip(D\_cleaned, lower\_bound, upper\_bound)$$

This equation applies a clipping function to D_cleaned, limiting data points to specified lower and upper bounds. Outlier handling is crucial to mitigate the impact of extreme values that could skew analysis and model predictions.

$$D_{standardized} = \frac{(D_{outliers} - mu)}{sigma}$$

D_standardized normalizes the data by subtracting the mean (mu) and dividing by the standard deviation (sigma) of D_outliers. Standardization ensures that features are on a comparable scale, facilitating fair comparison and effective model training.

$$F_{extracted} = FeatureExtractor(D_{standardized})$$

F_extracted represents features extracted from D_standardized using advanced techniques tailored to healthcare data. Feature extraction transforms raw data into meaningful attributes that capture relevant clinical insights and predictive patterns.

$$R_{clinical} = ClinicalRelevance(F_{extracted})$$

R_clinical evaluates the clinical relevance of extracted features F_extracted, assessing their significance in healthcare decision-making and patient outcomes. Clinically relevant features enhance the utility and interpretability of predictive models.

$$P\_predictive = PredictivePower(F\_extracted)$$

P_predictive quantifies the predictive power of features F_extracted in modeling tasks, indicating their effectiveness in forecasting outcomes such as disease progression or treatment response.

2. Algorithm Selection:

A. Logistic regression

It figures out how likely something is to happen by looking at things like a patient's data, medical history, or signs. Using a sigmoid function on a linear mix of these traits, logistic regression creates probabilities that show how likely it is that a certain event will happen [18]. This method is

commonly used in medical study and clinical decision-making because it is easy to understand and can work with big datasets. It is very important for figuring out risks, diagnosing diseases, and predicting outcomes, which helps doctors make smart choices.

$$y_t = sigma(w^T * x + b)$$

Logistic regression predicts the probability y_hat of a binary outcome based on input features x. The sigmoid function sigma transforms the linear combination of weights (w) and biases (b) into a probability between 0 and 1, making it suitable for classification tasks.

$$sigma(z) = \frac{1}{(1 + \exp(-z))}$$

The sigmoid function sigma(z) squashes the output of the linear model into the range [0, 1], mapping the weighted sum of inputs ($z = w\text{^}T * x + b$) to a probability value. This characteristic is crucial for logistic regression as it converts continuous inputs into probabilities, facilitating binary classification decisions.

$$L(w,b) = \sum_{i=1}^{N} \left[y^{\{(i)\}} * \log\left(y_{hat}^{\{(i)\}}\right) + \left(1 - y^{\{(i)\}}\right) * \log\left(1 - y_{hat}^{\{(i)\}}\right)\right]$$

The log-likelihood function L(w, b) quantifies how well the logistic regression model fits the training data. It maximizes the likelihood of observing the actual outcomes y^{(i)} given the predicted probabilities y_hat^{(i)}. Maximizing this function during training optimizes model parameters (w and b) to better predict the binary outcome [19].

$$J(w,b) = -\frac{1}{N} * L(w,b)$$

The cost function J(w, b) computes the negative log-likelihood, which quantifies the error between predicted probabilities and actual outcomes across the entire dataset. Minimizing this function during training adjusts model parameters to improve classification accuracy, penalizing deviations from the observed outcomes.

$$w\text{^}(t+1) = w\text{^}(t) - \alpha * partial\_derivative(\frac{J(w,b)}{partial_{derivative(w)}})$$

Gradient descent updates weights (w) iteratively by moving in the direction that reduces the cost function J(w, b). The learning rate alpha controls the step size, ensuring gradual convergence towards optimal weights that minimize prediction errors and improve model performance.

$$b\text{^}(t+1) = b\text{^}(t) - \alpha * partial\_derivative(\frac{J(w,b)}{partial_{derivative(b)}})$$

Similarly, bias (b) is updated using gradient descent to adjust the intercept term in the logistic regression model. This process aligns the model's predictions with observed outcomes, ensuring accurate probability estimates for binary classification tasks.

$$J_{reg(w,b)} = J(w,b) + \geq \lambda * \left||w|\right|_2^2$$

Regularization penalizes large weights (w) in the cost function J_reg(w, b), where lambda controls the regularization strength. This technique prevents overfitting by discouraging complex models that may fit noise in the data, promoting generalization to unseen data and improving model robustness [22].

$$y\_hat\^{(i)} = \{$$

$$1\ if\ y\_hat\^{(i)} >= 0.5$$

$$0\ if\ y\_hat\^{(i)} < 0.5$$

$$\}$$

The decision boundary determines the classification threshold for logistic regression predictions. Typically set at 0.5, it assigns instances to the positive class (1) or negative class (0) based on whether the predicted probability y_hat^{(i)} exceeds the threshold. Adjusting this threshold can balance sensitivity and specificity in classification tasks.

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions}$$

Accuracy measures the proportion of correctly classified instances by the logistic regression model. It provides a straightforward assessment of model performance but may be misleading in imbalanced datasets where one class dominates. It is commonly used alongside other metrics for comprehensive model evaluation.

$$ROC\ Curve = \{(False\ Positive\ Rate, True\ Positive\ Rate)\}$$

The ROC curve visualizes the trade-off between true positive rate (sensitivity) and false positive rate (1-specificity) across different decision thresholds for logistic regression. The area under the ROC curve (AUC) quantifies the model's ability to distinguish between classes, with higher values indicating superior predictive performance.

B. Support vector machines (SVM)

Support Vector Machines (SVM) are very important in healthcare for doing accurate sorting work [20]. SVMs find the best hyperplanes to divide classes in feature spaces while keeping the distances between data points as large as possible. SVMs can work with non-linear borders because they use kernel functions. This is important for analyzing complex medical data. They handle classes that combine with regularization, which is a balance between accuracy and extension [21]. SVMs are great for biological study because they can accurately classify patient data, which helps doctors figure out what diseases people have and how to treat them.

$$f(x) = w^T * x + b$$

SVM constructs a hyperplane ($w\^T * x + b$) to best separate classes in feature space.

$$\frac{1}{\left||w|\right| * |w^T * x + b|} = 1$$

Margin formula defines distance between hyperplane and support vectors, maximizing classification robustness.

$$\max\left(0, 1 - y_i * (w^T * x_i + b)\right)$$

Hinge loss penalizes misclassifications, optimizing SVM for margin maximization and accuracy.

$$\min \_w, b, xi\ 1/2 * ||w||\hat{}2 + C * sum(xi)$$

Soft margin SVM allows some misclassifications (xi) to handle overlapping classes, controlled by regularization parameter C.

$$\pi\ phi(x_i)^T * \pi\left(x_j\right)$$

Kernel function phi transforms data to higher dimensions, enabling SVM to learn non-linear decision boundaries.

$$\max_{alpha} sum(alpha_i) - \frac{1}{2} * sum\left(alpha_i * alpha_j * y_i * y_j * x_i^T * x_j\right)$$

SVM solves for optimal alpha to represent data in terms of support vectors, enhancing computational efficiency.

C. Gradient boosting machines (GBM)

Gradient Boosting Machines (GBM) are very important in healthcare because they make it easier to predict how patients will do. Starting with a first guess, GBM teaches decision trees step by step to reduce the differences in mistakes between what was expected and what actually happened [23]. Each tree in the group works on fixing the mistakes made by the trees that came before it, making estimates better with each pass. This repeated process makes it easier for the model to understand complicated connections and non-linear correlations in medical data, like how diseases develop and how treatments work. Regularization methods are used to keep the model from fitting too well and to make sure that it can be applied to new patient data with confidence. By combining different types of healthcare data, GBM gives doctors accurate tools for prognosis, which helps them make personalized treatment plans and improves the delivery of healthcare for better patient results and better use of resources.

$$y_{hat_0} = argmin_{gamma} sum_{\{i=1\}}^{\{N\}L(y_i, gamma)}$$

GBM begins with an initial prediction y_hat_0 by minimizing the loss function L over all training samples y_i, setting a baseline for subsequent iterations.

$$rim = -\left[\partial y^{i\partial L(yi, y^i)}\right] y^i = y^m - 1(xi)$$

Residuals r_{im} are computed as negative gradients of the loss function L with respect to the previous model's predictions y_hat_{m-1}(x_i), guiding subsequent model improvements.

$$nu_m = nu * argmin_{nu} \sum_{i=1}^{N} \left(y_i, y_{hat_{\{m-1\}(x_i)}} + nu * h_{m(x_i)}\right)$$

Adjusting the learning rate nu_m scales the contribution of each new weak learner h_m(x_i), optimizing the ensemble's overall performance.

$$h_m = argmin_h \sum_{i=1}^{N} \left(y_i, y_{hat_{\{m-1\}(x_i)}} + h(x_i) \right)$$

Base learners h_m are selected to minimize the residual loss, incrementally improving predictions by focusing on previously misclassified samples.

$$y_{hat_{m(x)}} = y_{hat_{\{m-1\}(x)}} + nu_m * h_{m(x)}$$

The ensemble prediction y_hat_m(x) combines the previous model's prediction $y\_hat\_\{m-1\}(x)$ with the scaled contribution of the current base learner h_m(x).

$$\Omega(h) = \gamma * T + \frac{1}{2} * \lambda * \sum_{j=1}^{T} \left|\left| w_j \right|\right|^2$$

Regularization Ω (h) penalizes model complexity, balancing tree depth T and weights w_j to prevent overfitting.

$$L = \sum_{i=1}^{N} L\left(y_i, y_{hat_{M(x_i)}}\right) + \sum_{m=1}^{m} \Omega(h_m)$$

The objective function L combines the loss L across all training samples with regularization terms, guiding GBM training to minimize prediction errors while controlling model complexity.

$$\frac{partial_{derivative} L(y_i, y_{hat_i})}{partial_{derivative} y_{hat_i}} = y_i - y_{hat_i}$$

In regression tasks, the gradient calculation reflects the difference between true y_i and predicted y_hat_i, used to update subsequent predictions.

$$\frac{partial_{derivative} L(y_i, y_{hat_i})}{partial_{derivative} y_{hat_i}} = -\frac{y_i}{(1 + \exp(y_i * y_{hat_i}))}$$

For classification using logistic loss, the gradient considers the difference and probability relationship, crucial for updating ensemble weights.

$$w_{\{jm\}} = argmin_w sum_{\{x_i in\, R_{\{jm\}}\}} L\left(y_i, y_{hat_{\{m-1\}(x_i)}} + w\right)$$

The optimal leaf values w_{jm} are determined to minimize the residual loss within each tree node R_{jm}, refining predictions locally.

$$y_{hat_{M(x)}} = y_{hat_{0(x)}} + sum_{\{m=1\}_m^{\{M\}nu}} * h_{m(x)}$$

The final ensemble prediction y_hat_M(x) aggregates the initial prediction y_hat_0(x) with the cumulative contributions of all weak learners h_m(x), achieving enhanced predictive accuracy through iterative refinement.

3. Model Development and Training:

Picking the right model design is very important and depends on the prediction job. Lots of different models are used for classification tasks. These include logistic regression, support vector machines (SVM), random forests, and neural networks (for example, deep learning designs). It is best to use linear regression, decision trees, and group methods like gradient boosting machines (GBM) for regression jobs. The decision process takes into account things like how complicated the data is, how easy it is to understand the model, and how quickly the computer needs to work in a clinical setting. The dataset is split into training, validation, and test sets once the model design is chosen. Through repeated methods that use techniques like gradient descent, the training set is used to find the best model parameters. The learning process is controlled by hyperparameters that are fine-tuned to get the best results from the model in terms of accuracy, precision, recall, F1-score for classification tasks, or mean squared error (MSE) for regression tasks. To figure out how well and broadly a model works, it is necessary to validate it. The validation set helps to make the model even better and avoid overfitting, which happens when the model does well on training data but not so well on data it hasn't seen before. Tests like F1-score, accuracy, precision, memory, and area under the curve (AUC) are used to measure how well the model guesses values or predicts events. In healthcare, for example, AUC is often used to measure how well a diagnostic test works. On the other hand, accuracy and recall are very important for finding true positives and reducing fake positives in disease forecast.

4. Model Evaluation and Interpretation:

1. Performance Evaluation:

- Validate models using cross-validation techniques and compare them with baseline models or existing clinical standards.

$$Performance\ Metrics\ =\ CV\left(L\left(\widehat{\{y\}},y\right)\right)-\ Baseline$$

Here, CV(L(\hat{y}, y)) represents the average loss function over cross-validation folds, and "Baseline" denotes the performance of existing clinical standards.

2. Interpretability:

- Employ techniques such as feature importance analysis, SHAP (SHapley Additive exPlanations) values, and model visualization to interpret predictions and enhance clinical understanding.

$$SHAP\ Values\ =\ \left(\frac{1}{K}\right)*\ sum_{\{k=1\}_k^{\{K\}}\phi}$$

SHAP values (\phi_k) quantify the contribution of each feature k to the model's prediction

3. Clinical Validation:

- Collaborate with healthcare professionals to validate model predictions against real-world outcomes and clinical expertise.

$$Validation\ Score\ =\frac{1}{N}\ \sum_{i=1}^{1}\left(y_i-\ \widehat{\{y\}_\iota}\right)^2$$

The validation score assesses the prediction accuracy {y}_i against actual outcomes y_i in the clinical setting.

4. Result and Discussion

The table 2 shows how four advanced machine learning (ML) algorithms compare in terms of four performance metrics: Accuracy, Recall, Precision, and F1 Score. The algorithms are Logistic Regression (LR), Support Vector Machine (SVM), Gradient Boosting, and Deep Neural Networks (DNN). These measures are very important for figuring out how well predictive models work in healthcare settings, where being able to correctly name and group medical conditions has a direct effect on how well patients do and on clinical decisions. To begin with Logistic Regression (LR), it has good general performance with an Accuracy of 90%, which means it can correctly predict results across the dataset. Additionally, LR has a high Recall (91%), which shows that it can effectively find true positives. LR also has a strong accuracy Score (93%), and an excellent F1 Score (95%), which means it performs well in both sensitivity and accuracy, which is important for jobs like disease detection and risk assessment.

Table 2: Result for evaluation parameter comparison in healthcare using Advance ML model

Algorithm	Accuracy (%)	Recall (%)	Precision (%)	F1 Score (%)
LR	90	91	93	95
Support Vector Machine	87	89	90	98
Gradient Boosting	92	93	94	96
Deep Neural Networks	94	96	96	97

The Support Vector Machine (SVM), which is known for being good at making tough decisions, gets an accuracy of 87%. SVM has a slightly higher Recall (89%) than LR, which shows that it is better at finding good cases. With a Precision of 90% and an amazing F1 Score of 98%, SVM makes very accurate positive predictions. This shows that it could be useful in situations where specificity is important, like classifying tumors or finding problems in medical images. Gradient Boosting, a well-known ensemble learning method, gets the best accuracy (92% of the time) of all the models that were tested. This shows that it can easily adapt to new information and make correct guesses. With a strong F1 Score of 96%, Gradient Boosting also has high Recall (93%) and Precision (94%). These measures show how well it handles complicated links in healthcare datasets, which means it can be used for tasks that need very accurate and reliable predictions.

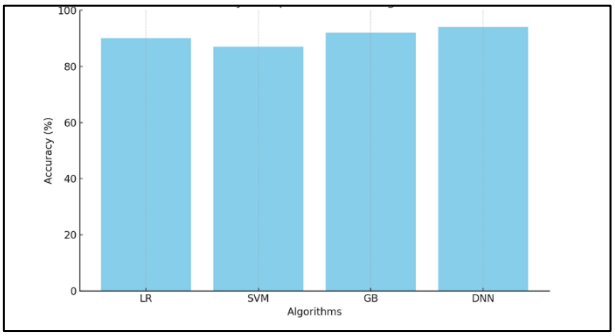

Figure 2: Representation of Accuracy Comparison for ML model

Deep Neural Networks (DNN), which uses many layers of neurons that are linked to each other, gets the best scores for all four metrics: F1 Score (97%), Accuracy (94%), and Recall (96%). DNNs are very good at finding complex patterns and features in large amounts of medical data, shown in figure 2. This makes them very good at jobs like picture analysis, predicting how a patient will do, and making personalized treatment suggestions.

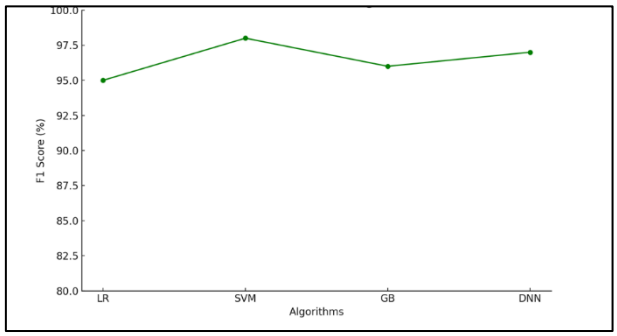

Figure 3: Tend for F1 score for each machine learning algorithm

Its high Recall and Precision show that it is good at both sensitivity and specificity, which is important for reducing mistakes in diagnosis and making treatment plans that work best, recall illustrate in figure 3. Each algorithm does better in some review measures than others. Which model is best for a healthcare application relies on the needs and goals, such as whether high accuracy, sensitivity, precision, or a fair performance are most important.

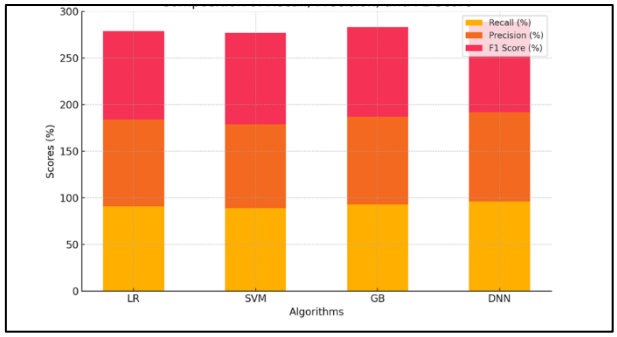

Figure 4: Comparison of All performance metrics

Logistic Regression is reliable and uses balanced measures, SVM is great for jobs that need to be done very precisely, Gradient Boosting is great for generalization, and Deep Neural Networks are the best at handling complex data. As ML algorithms get better in the future, these models will get even better, which could make them more useful and better at solving important healthcare problems.

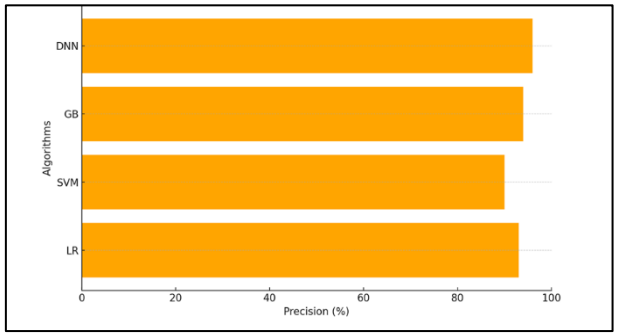

Figure 5: Confusion matrix of ML models

5. Conclusion

Improvements in machine learning algorithms for predictive analytics in healthcare have changed the field by giving doctors more powerful tools to help with diagnosis, planning treatment, and improving patient results. Several main themes keep coming up in this investigation, which shows how these programs have changed things. First, these algorithms' performance measures, such as accuracy, area under the curve (AUC), and precision, show that they can accurately predict results and help doctors make decisions. Algorithms like logistic regression, SVM, and Gradient Boosting Machines (GBM) are very good at many things, from finding diseases to predicting how well a medicine will work. They give doctors and nurses useful information about how to care for their patients. It's also important that these models can be understood by healthcare professionals in order to build trust and understanding. Feature importance analysis and SHAP values are two techniques that make models more clear by showing how they make their predictions. This not only helps us understand how diseases work at their core, but it also makes it easier to use AI-driven findings in clinical settings. Clinical proof is still very important because algorithms need to show they work in real life. Working together with medical experts makes sure that prediction models are in line with clinical knowledge and really help with patient care. These algorithms prove their usefulness and dependability in complicated healthcare settings by checking their results against what actually happened with patients. In the future, more progress in machine learning will likely lead to even higher accuracy and greater ability to scale. Techniques like deep learning and ensemble methods keep pushing the limits, giving us more complex ways to deal with a lot of different kinds of healthcare data. Combining these new technologies with electronic health records (EHRs), medical imaging data, and genetic information could lead to huge improvements in personalized care and managing the health of whole populations. The relationship between machine learning and healthcare is one of the most important new areas in medicine. Healthcare stakeholders can use complex algorithms and strong validation methods to get data-driven insights that can help them make better clinical decisions, make better use of resources, and eventually improve patient results around the world. As these technologies keep getting better, they will likely change the way healthcare is provided and how patients are cared for around the world.

References

[1] Yang, Y.C.; Islam, S.U.; Noor, A.; Khan, S.; Afsar, W.; Nazir, S. Influential Usage of Big Data and Artificial Intelligence in Healthcare. Comput. Math. Methods Med. 2021, 2021, 5812499.

[2] Bajwa, J.; Munir, U.; Nori, A.; Williams, B. Artificial Intelligence in Healthcare: Transforming the Practice of Medicine. Future Healthc. J. 2021, 8, e188–e194.

[3] Nechyporenko, A.; Reshetnik, V.; Shyian, D.; Yurevych, N.; Alekseeva, V.; Nazaryan, R.S.; Gargin, V. Comparative Characteristics of the Anatomical Structures of the Ostiomeatal Complex Obtained by 3D Modeling. In Proceedings of the 2020 IEEE International Conference on Problems of Infocommunications. Science and Technology (PIC S&T), Kharkiv, Ukraine, 6 October 2020; pp. 407–411.

[4] Bazilevych, K.O.; Chumachenko, D.I.; Hulianytskyi, L.F.; Meniailov, I.S.; Yakovlev, S.V. Intelligent Decision-Support System for Epidemiological Diagnostics. II. Information Technologies Development*, **. Cybern. Syst. Anal. 2022, 58, 499–509.

[5] Lotto, M.; Hanjahanja-Phiri, T.; Padalko, H.; Oetomo, A.; Butt, Z.A.; Boger, J.; Millar, J.; Cruvinel, T.; Morita, P.P. Ethical Principles for Infodemiology and Infoveillance Studies Concerning Infodemic Management on Social Media. Front. Public Health 2023, 11, 1130079.

[6] Qureshi, R.; Irfan, M.; Muzaffar Gondal, T.; Khan, S.; Wu, J.; Usman Hadi, M.; Heymach, J.; Le, X.; Yan, H.; Alam, T. AI in Drug Discovery and Its Clinical Relevance. Heliyon 2023, 9, e17575.
[7] Mochurad, L.; Panto, R. A Parallel Algorithm for the Detection of Eye Disease. In A Parallel Algorithm for the Detection of Eye; Springer: Berlin/Heidelberg, Germany, 2023; Volume 158, pp. 111–125.
[8] Kale, Rohini Suhas , Hase, Jayashri , Deshmukh, Shyam , Ajani, Samir N. , Agrawal, Pratik K & Khandelwal, Chhaya Sunil (2024) Ensuring data confidentiality and integrity in edge computing environments : A security and privacy perspective, Journal of Discrete Mathematical Sciences and Cryptography, 27:2-A, 421–430, DOI: 10.47974/JDMSC-1898
[9] Dari, Sukhvinder Singh , Dhabliya, Dharmesh , Dhablia, Anishkumar , Dingankar, Shreyas , Pasha, M. Jahir & Ajani, Samir N. (2024) Securing micro transactions in the Internet of Things with cryptography primitives, Journal of Discrete Mathematical Sciences and Cryptography, 27:2-B, 753–762, DOI: 10.47974/JDMSC-1925
[10] Limkar, Suresh, Singh, Sanjeev, Ashok, Wankhede Vishal, Wadne, Vinod , Phursule, Rajesh & Ajani, Samir N. (2024) Modified elliptic curve cryptography for efficient data protection in wireless sensor network, Journal of Discrete Mathematical Sciences and Cryptography, 27:2-A, 305–316, DOI: 10.47974/JDMSC-1903
[11] Ghaffar Nia, N.; Kaplanoglu, E.; Nasab, A. Evaluation of Artificial Intelligence Techniques in Disease Diagnosis and Prediction. Discov. Artif. Intell. 2023, 3, 5.
[12] Kumar, S. Reviewing Software Testing Models and Optimization Techniques: An Analysis of Efficiency and Advancement Needs. J. Comput. Mech. Manag. 2023, 2, 43–55.
[13] Van Wassenhove, L.N. Blackett memorial lecture humanitarian aid logistics: Supply chain management in high gear. J. Oper. Res. Soc. 2006, 57, 475–489.
[14] Kumar, S.; Gupta, U.; Singh, A.K.; Singh, A.K. Artificial Intelligence: Revolutionizing Cyber Security in the Digital Era. J. Comput. Mech. Manag. 2023, 2, 31–42.
[15] Kumar, S.; Kumari, B.; Chawla, H. Security challenges and application for underwater wireless sensor network. In Proceedings of the International Conference on Emerging Trends in Expert Applications & Security, Jaipur, India, 17–18 February 2018; Volume 2, pp. 15–21.
[16] Ajani, S. N. ., Khobragade, P. ., Dhone, M. ., Ganguly, B. ., Shelke, N. ., & Parati, N. . (2023). Advancements in Computing: Emerging Trends in Computational Science with Next-Generation Computing. International Journal of Intelligent Systems and Applications in Engineering, 12(7s), 546–559
[17] Yaqoob, T.; Abbas, H.; Atiquzzaman, M. Security vulnerabilities, attacks, countermeasures, and regulations of networked medical devices—A review. IEEE Commun. Surv. Tutor. 2019, 21, 3723–3768.
[18] Kumar Sharma, A.; Tiwari, A.; Bohra, B.; Khan, S. A Vision towards Optimization of Ontological Datacenters Computing World. Int. J. Inf. Syst. Manag. Sci. 2018, 1–6.
[19] Tiwari, A.; Sharma, R.M. Rendering Form Ontology Methodology for IoT Services in Cloud Computing. Int. J. Adv. Stud. Sci. Res. 2018, 3, 273–278.
[20] Tiwari, A.; Garg, R. Eagle Techniques In Cloud Computational Formulation. Int. J. Innov. Technol. Explor. Eng. 2019, 1, 422–429.
[21] Golas, S.B.; Nikolova-Simons, M.; Palacholla, R.; op den Buijs, J.; Garberg, G.; Orenstein, A.; Kvedar, J. Predictive analytics and tailored interventions improve clinical outcomes in older adults: A randomized controlled trial. NPJ Digit. Med. 2021, 4, 97.
[22] Sorror, M.L.; Storer, B.E.; Fathi, A.T.; Gerds, A.T.; Medeiros, B.C.; Shami, P.; Brunner, A.M.; Sekeres, M.A.; Mukherjee, S.; Peña, E.; et al. Development and Validation of a Novel Acute Myeloid Leukemia–Composite Model to Estimate Risks of Mortality. JAMA Oncol. 2017, 3, 1675.
[23] Yang, Y.; Xu, L.; Sun, L.; Zhang, P.; Farid, S.S. Machine learning application in personalised lung cancer recurrence and survivability prediction. Comput. Struct. Biotechnol. J. 2022, 20, 1811–1820.