International Journal of Emerging Research in Engineering and Technology
Pearl Blue Research Group | Volume 7 Issue 3 PP 1-9, 2026
ISSN: 3050-922X | https://doi.org/10.63282/3050-922X.IJERET-V7I3P101

Original Article

Machine Learning Based Clinical Trial Performance Prediction for Medical Industry Applications

Dr. Nilesh Jain
Associate Professor, Department of Computer Sciences and Applications, Mandsaur University, Mandsaur, India.

Received On: 16/05/2026 ***Revised On: 10/06/2026*** ***Accepted On: 21/06/2026*** ***Published On: 02/07/2026***

Abstract - Clinical trials are essential for evaluating the safety and effectiveness of medical treatments; however, they often involve high costs, long durations, and uncertain outcomes. This study investigates the possibility of predicting clinical trial results using machine learning (ML) techniques by analyzing patients' intraoperative and postoperative adverse reactions, postoperative vital signs, and satisfaction levels in trials of new sedative drugs. A Clinical Trials dataset comprising 13,748 records and 11 attributes was utilized, followed by comprehensive preprocessing steps including missing value handling, label encoding, feature scaling, and data balancing using SMOTE to address class imbalance. Machine learning models including Logistic Regression (LR), Random Forest (RF), Convolutional Neural Networks (CNN), and the proposed Extreme Gradient Boosting (XGBoost) were implemented for comparative analysis. Experimental results demonstrate that XGBoost outperforms all other models with an accuracy of 92.7%, precision of 95.6%, recall of 95.9%, and F1-score of 95.7%, indicating superior predictive capability and robustness. The comparative analysis confirms that ensemble-based learning methods, particularly XGBoost, are highly effective for clinical trial outcome prediction, offering improved reliability and decision-support potential in medical research.

Keywords - Machine Learning, Clinical Decision Support, Risk Stratification, Deep Learning, Electronic Health Records, Artificial Intelligence, Predictive Modeling.

1. Introduction

Clinical trials play a vital role in the development and evaluation of new drugs, medical devices, and therapeutic interventions. However, the increasing complexity of modern clinical trials has led to higher costs, longer durations, and greater risks of failure. Contemporary trials involve multiple endpoints, extensive eligibility criteria, numerous study sites, and large patient populations, making trial management and performance prediction more challenging than ever before [1][2][3]. Accurate prediction of clinical trial outcomes is therefore essential for improving operational efficiency, reducing costs, and enhancing decision-making throughout the clinical research process. The growing availability of healthcare data has created new opportunities to apply advanced computational techniques for predicting and optimizing clinical trial performance[4][5][6].

The healthcare sector has witnessed an unprecedented growth in data generated from Electronic Health Records (EHRs), laboratory tests, wearable devices, medical imaging systems, and patient-reported outcomes. Traditional statistical and rule-based approaches are often limited in their ability to analyze such large and heterogeneous datasets[7][8]. Consequently, there is an increasing demand for intelligent analytical techniques capable of extracting meaningful information from complex healthcare data. Predictive analytics has emerged as a valuable approach for identifying patterns and trends that can support evidence-based decision-making and improve healthcare outcomes [9][10][11].

Clinical Prediction Models (CPMs) have become important tools for estimating disease risk, treatment response, patient prognosis, and healthcare resource utilization. These models integrate multidimensional patient information to support clinical decision-making and improve the quality and efficiency of healthcare services[12][13]. The growing adoption of digital health technologies and advanced computational methods has further enhanced the development of predictive models capable of handling large-scale medical data. Such models can assist healthcare professionals and clinical researchers in identifying potential risks, improving trial management, and making more informed decisions throughout the clinical research lifecycle [14][15][16][17].

Recent advances in Artificial Intelligence (AI), particularly Machine Learning (ML) and Deep Learning (DL), have created new opportunities for predicting and optimizing clinical trial performance. ML algorithms can automatically learn patterns from historical clinical data and generate accurate predictions regarding trial success, patient enrollment, treatment outcomes, and operational risks. Deep Learning, a specialized subset of ML, utilizes multi-layer neural networks to analyze complex structured and unstructured data, including medical images, clinical notes, and genomic information[18]. By leveraging these advanced techniques, researchers can reduce trial complexity, improve efficiency, minimize failure risks, and accelerate the development of effective medical treatments. As a result, ML- and DL-based predictive systems are becoming increasingly important tools for enhancing clinical trial performance and supporting data-driven decision-making in medical industry applications[19][20][21].

1.1. Motivation and Contribution

The motivation for this study is the growing complexity, cost, and unpredictability of clinical trials in the modern era, causing delays and high failure rates in drug development processes. The established analytical techniques fall short of being able to effectively represent the intricate relationships found in clinical trial data, particularly with large-scale and heterogenous healthcare data. As the amount of electronic health records and clinical datasets is rapidly increasing, the demand for intelligent and data-driven methods that would be accurate to foresee clinical trial outcomes is a burning topic. The potential for machine learning to learn the hidden patterns in historical data and to enhance the efficiency of decision making makes it an ideal candidate for developing a strong predictive model for clinical trial performance analysis. This research offers several key contributions as listed below:

- Utilized a large-scale Clinical Trials dataset with 13,748 records and 11 attributes for robust analysis.
- Analyzed clinical trial performance data to understand patient outcomes, including intraoperative and postoperative adverse reactions, vital signs, and satisfaction levels.
- Developed a machine learning-based framework for predicting clinical trial performance in sedative drug studies.
- Focused on improving prediction reliability for clinical trial success and failure classification.
- Proposed an Extreme Gradient Boosting (XGBoost) based model for accurate and efficient clinical trial performance prediction.
- Provided a reliable decision-support approach for improving clinical trial analysis and medical research efficiency.

The ***justification*** of this study lies in the growing need to improve the efficiency, accuracy, and reliability of clinical trial outcome prediction, as traditional statistical methods and basic machine learning models often fail to capture complex relationships in clinical data. The proposed framework uses advanced preprocessing techniques like SMOTE-based class balancing and feature standardization to ensure high-quality input data for model training, resulting in more reliable predictions. The ***novelty*** of this work lies in the application of as well as comparative evaluation of several machine learning and deep learning models for clinical trial performance prediction and integrating an optimized structured clinical data machine learning model, Extreme Gradient Boosting (XGBoost). This has not only improved the ability to predict the results of these trials but also offers a strong and scalable decision support framework for the analysis of sedative drug trial results.

1.2. Organization of the Paper

The paper is organized as follows: **Section II** presents the related work on clinical trial performance prediction, **Section III** describes the dataset, preprocessing techniques, and proposed model implementation, **Section IV** discusses the experimental results and comparative performance analysis of different models, and **Section V** concludes the study with key findings and future research directions.

2. Literature review

A comprehensive review of recent studies on Clinical Trial Performance Prediction was conducted to support the development of this research. Table I summarizes the existing literature by highlighting the proposed models, datasets used, key findings, and challenges identified in previous studies.

Upadhaya *et al.*, (2026) explored a multi-regional imaging-omics approach using CT-based radiomics and 3D dosiomics features to predict radiation pneumonitis (RP) in 451 patients with locally advanced non-small cell lung cancer (LA-NSCLC) from the NRG Oncology/RTOG 0617 trial. Features were extracted from CT images and 3D dose maps using multi-regional lung contours. LASSO-GLM, Random Forest, XGBoost, and SVM (linear and RBF) models were evaluated. The combined radiomics and dosiomics model from total lungs achieved the highest performance with an AUC of 0.74 (95% CI, 0.65–0.83) on the validation dataset and 0.70 (95% CI, 0.56–0.83) on the test dataset using SVM-RBF. Combining radiomics, dosiomics, and filter-based features from the contralateral lung also achieved an AUC of 0.69 on the test dataset using SVM-RBF, similar to total-lung dosiomics features[22].

Baihaqi and Kurniawan, (2025) The study developed a robust predictive model for clinical toxicity classification using the Clintox dataset containing 1,431 chemical compounds. The methodology incorporated feature selection to remove irrelevant attributes and hyperparameter tuning to optimize performance. The proposed Cuckoo Search-Ensemble model demonstrated improved predictive capability and robustness against data variations. Model performance was evaluated using accuracy, precision, recall, and F1-score. The best results were achieved by the AdaBoost model with a learning rate of 0.1 and 150 estimators, attaining an F1-score of 0.6515 and an accuracy of 0.9125[23].

Tasnim, Hossain and Hasan, (2025) The study developed and validated machine learning models to predict treatment response and optimize patient outcomes in clinical trials using data from 2,156 patients enrolled in phase II–III oncology trials (2018–2024). Logistic regression, Random Forest, SVM, XGBoost, and ensemble models were evaluated using clinical, molecular, and laboratory features, with external validation on 856 independent patients. The ensemble model achieved the best performance with an AUC-ROC of 0.862 (95% CI: 0.837–0.887), outperforming traditional clinical scoring (AUC-ROC: 0.698, $p<0.001$), and maintained robust external validation performance (AUC-ROC: 0.831). Mutation burden and performance status were the most predictive features. The adaptive randomization algorithm allocated 68.4% of patients to optimal treatments compared to 50% with conventional randomization, improving overall response rates by 12.3% ($p<0.001$), demonstrating the clinical utility of machine learning for treatment optimization and patient stratification in precision oncology[24].

Devi *et al.*, (2024) examined the importance of predictive models in estimating the probability of achieving clinical trial enrollment goals before trial initiation. A novel prediction

framework was developed using data-driven approaches and statistical analyses. GloVe and BioBERT embeddings were used to convert textual inputs into vector representations, which were combined with the top 11 features selected through correlation analysis and data from ClinicalTrials.gov. The proposed MIRLR ensemble stacking method integrated M5P, IBK, Random Forest, Logistic Regression, and REPTree models, with their predictions used as input to a final Random Tree Regressor (RTR) for enrollment rate forecasting. The MIRLR model achieved a root mean square error (RMSE) of 0.334 patients/day on the test dataset and 0.0057 patients/day on the training dataset using a 70:30 data split, demonstrating strong predictive performance in forecasting clinical trial site enrollment rates[25].

Reinisch *et al.*, (2024) proposed CTP-LLM, the first Large Language Model (LLM)-based approach for Clinical Trial Outcome Prediction (CTOP), and introduced the Phase Transition (PT) Dataset, which labels trials according to their progression through the regulatory process. The fine-tuned GPT-3.5-based CTP-LLM predicts clinical trial phase transitions directly from original protocol texts without requiring manually selected features. The model achieved an accuracy of 67% in predicting phase transitions across all trial phases and 75% accuracy in predicting transitions from Phase III to final approval, demonstrating the potential of LLM-based methods for forecasting clinical trial outcomes and evaluating trial design[26].

Lu *et al.*, (2022) examined long-term mortality in the National Lung Screening Trial (NLST) using a deep learning model based on 3D-ResNet for predicting non-lung-cancer mortality (cardiovascular and respiratory mortality). Using 3D CT scans and clinical information, the model achieved an AUC of 0.73, an F1-score of 0.60, and a Matthews Correlation Coefficient of 0.38, outperforming human experts in cardiovascular mortality prediction. Saliency map analysis identified thoracic regions associated with mortality risk, supporting earlier detection and targeted preventive interventions[27].

Table 1: Recent Studies on Clinical Trial Performance Prediction Using Machine Learning Technqiues

Author	Proposed Work	Dataset	Results	Limitations & Future Work
Upadhaya et al. (2026)	Developed a multi-regional imaging-omics framework using radiomics and dosiomics features to predict radiation pneumonitis (RP) in LA-NSCLC patients. Evaluated LASSO-GLM, RF, XGBoost, and SVM models.	NRG Oncology/RTOG 0617 Clinical Trial Dataset (451 patients)	Best performance achieved using SVM-RBF with radiomics and dosiomics features, obtaining AUC = 0.74 on validation data and AUC = 0.70 on test data.	Performance remains moderate; future work should incorporate larger multi-center datasets and additional clinical biomarkers to improve prediction accuracy.
Baihaqi and Kurniawan (2025)	Proposed a Cuckoo Search-Ensemble framework with feature selection and hyperparameter tuning for clinical toxicity prediction.	Clintox Dataset (1,431 chemical compounds)	AdaBoost achieved 91.25% accuracy and F1-score of 0.6515.	Limited dataset size may affect generalization. Future studies should evaluate larger toxicity datasets and advanced ensemble methods.
Tasnim, Hossain and Hasan (2025)	Developed machine learning and adaptive randomization models for treatment response prediction in oncology clinical trials.	Phase II–III Oncology Trial Dataset (2,156 patients + 856 external validation patients)	Ensemble model achieved AUC-ROC = 0.862 and 0.831 on external validation. Adaptive randomization improved response rates by 12.3%.	Study focused on oncology trials only. Future work should validate the framework across multiple disease domains.
Devi et al. (2024)	Proposed MIRLR, an ensemble stacking framework combining M5P, IBK, RF, LR, and REPTree models for enrollment rate prediction.	ClinicalTrials.gov Dataset	Achieved RMSE = 0.334 patients/day on testing data and 0.0057 patients/day on training data.	Performance depends on selected features and trial characteristics. Future work may explore deep learning and real-time enrollment prediction.
Reinisch et al. (2024)	Introduced CTP-LLM, a GPT-3.5-based large language model for	Phase Transition (PT) Dataset	Achieved 67% accuracy across all phases and 75% accuracy for Phase	Accuracy remains limited for some trial phases. Future work should

	clinical trial phase transition prediction.		III to approval prediction.	investigate larger LLMs and multimodal clinical trial information.
Lu et al. (2022)	Applied a deep learning-based 3D-ResNet model for long-term mortality prediction using CT scans and clinical information.	National Lung Screening Trial (NLST) Dataset	Achieved AUC = 0.73, F1-score = 0.60, and MCC = 0.38, outperforming human experts in cardiovascular mortality prediction.	Limited interpretability and moderate predictive performance. Future work should integrate explainable AI techniques and additional clinical variables.

Research gaps: Despite significant advancements in machine learning and deep learning techniques for clinical trial outcome prediction, several research gaps still exist. To the best of our knowledge, most of the existing studies are focused on, e.g., toxicity prediction, mortality analysis, or forecasting phase transitions, while few studies have focused on the prediction of the overall performance of clinical trials across a range of diseases involving a combination of patient-level and trial-level features. In addition, many approaches only use imaging data or structured clinical data, but not combine heterogeneous features to get a better predictive accuracy. Class imbalance, data quality issues, and lack of standardized preprocessing pipelines further limit model generalization in real-world clinical trial settings. Therefore, there is a need for a robust and scalable machine learning framework that integrates effective preprocessing, handles data imbalance, and provides a comprehensive comparative analysis to improve prediction reliability and decision-making in clinical trials.

3. Research Methodology

The proposed methodology involves collecting a Clinical Trials dataset containing 13,748 records with 11 attributes, followed by exploratory data analysis using bar plots and heatmaps to understand data distributions and relationships. Data preprocessing is performed by handling missing values, removing duplicates, applying label encoding, and normalizing features using StandardScaler to improve data quality. To address class imbalance, the SMOTE technique is applied to generate synthetic samples for the minority class, and the dataset is split into training and testing sets using a 70:30 stratified approach. Finally, an XGBoost-based supervised learning model is developed for clinical trial performance prediction and evaluated using accuracy, precision, recall, F1-score, and AUC-ROC metrics. **Fig. 1** illustrates the proposed flowchart for Clinical Trial Performance Prediction using machine learning.

The following section presents a detailed description of each step involved in the proposed methodology:

3.1. Data Gathering and Analysis

The Clinical Trials dataset was collected from publicly available clinical trial records and contains information on trial phases, recruitment status, interventions, and study outcomes. The dataset comprises 13,748 records and 11 attributes, providing a comprehensive foundation for clinical trial performance prediction. To explore data characteristics and identify underlying patterns, visualization techniques such as bar plots and heatmaps were employed to analyze feature distributions and relationships.

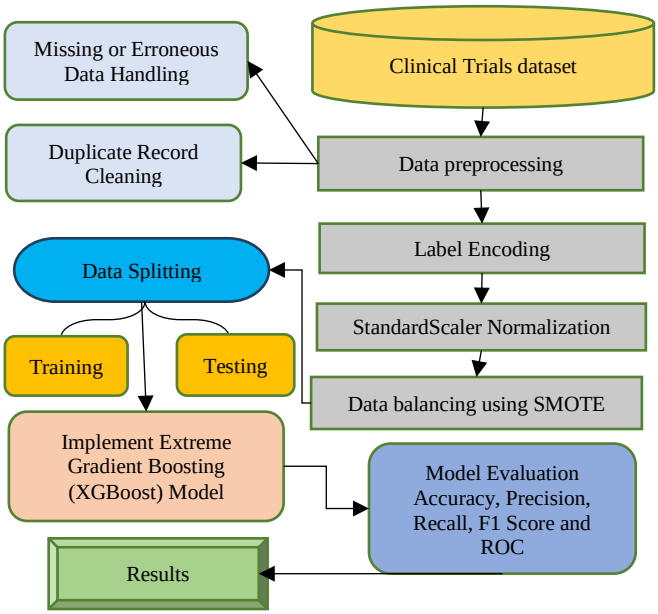

Fig 1: Proposed flowchart for Clinical Trial Performance Prediction using machine learning

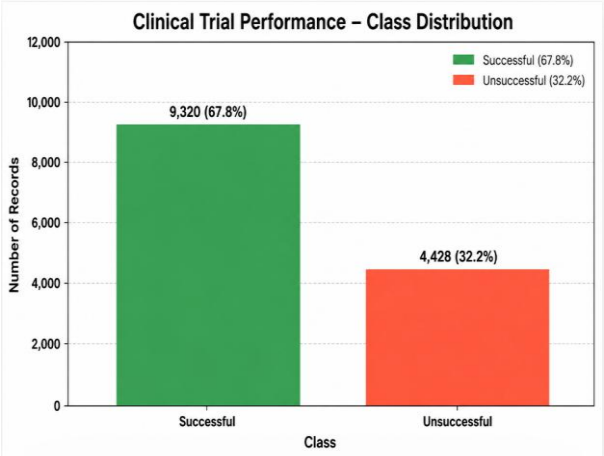

Fig 2: Bar Graph of Class Distribution

Fig. 2 shows a bar graph of the clinical trial performance classes of the data set. The largest number of records are classified in the Successful category (67.8%) and the Unsuccessful category (32.2%). The distribution shows a moderate level of class imbalance, suggesting that successful trials are more likely to happen than unsuccessful ones. The visualization provides a clear overview of the target variable distribution and highlights the need for appropriate data balancing techniques to ensure unbiased and reliable performance of machine learning models during clinical trial outcome prediction.

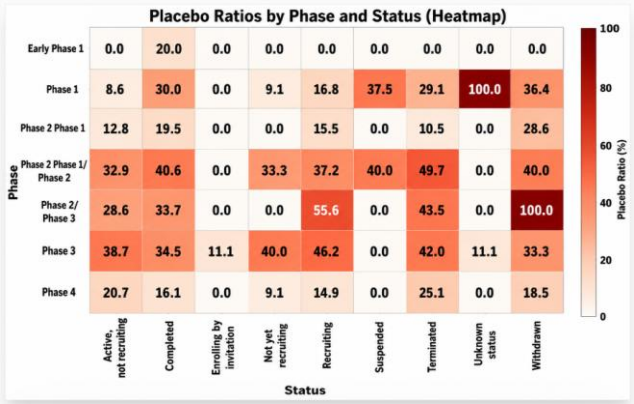

Fig 3: Correlation Matrix Heatmap

Figure 3 illustrates the distribution of placebo ratios across different clinical trial phases and trial statuses. Darker shades of red indicate higher placebo ratios, while lighter shades represent lower values. The figure reveals that placebo usage varies considerably across phases and statuses, with the highest ratios observed in the Unknown Status category for Phase 1 and the Withdrawn category for Phase 2/Phase 3, both reaching 100%. Moderate placebo ratios are evident in recruiting, suspended, terminated, and completed trials, particularly in advanced phases. Overall, the heatmap provides a clear visual representation of the relationship between clinical trial phase, trial status, and placebo utilization, helping to identify patterns and trends within the clinical trial dataset.

3.2. Data Pre-processing

The Clinical Trials dataset underwent a structured data preparation process that included data cleaning and integration. Preprocessing steps involved handling missing and erroneous values, removing duplicate records, encoding categorical features, and normalizing data values. These procedures improved data quality and enhanced the predictive performance of the proposed machine learning model:

- **Missing or Erroneous Data Handling:** Missing or erroneous data refer to incomplete, inaccurate, or inconsistent values within a dataset, which can negatively impact the performance and reliability of machine learning models. Therefore, such issues must be addressed through appropriate data preprocessing techniques, including imputation, removal, or correction of invalid values, to ensure data quality and improve model accuracy.

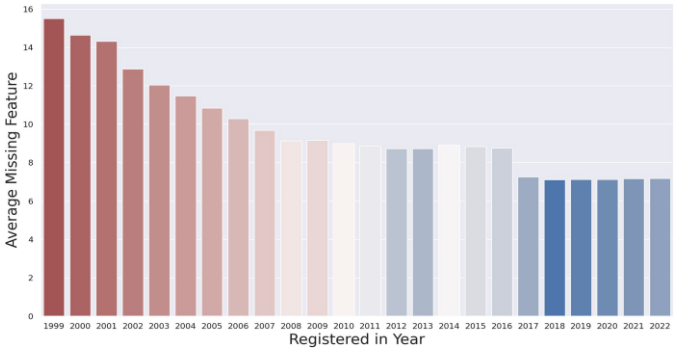

Fig 4: Average Missing Value Rates

Figure 4 illustrates the trend in the average number of missing features in clinical trial data over time. The results show a gradual decline in missing values from earlier years to more recent years, indicating continuous improvements in data quality, reporting practices, and record completeness. This change is indicative of better data management standards and increased uniformity with clinical trial documentation throughout time.

- **Duplicate Record Cleaning:** The dataset was divided into training and testing sets using a stratified sampling approach with a 70:30 ratio. This method ensured that the class distribution in both subsets remained consistent with the original dataset, thereby preserving proportional representation and improving the reliability and generalization performance of the model.
- **Label Encoding:** Categorical features, such as trial phases, recruitment status, and outcome labels, were transformed into numerical representations using label encoding to ensure compatibility with machine learning algorithms and facilitate effective model training.

3.3. StandardScaler Normalization

Given the different scales of each descriptor, the dataset was standardized using the StandardScaler () method to transform the data so that the mean of the resulting distribution is zero and the standard deviation is one. This transformation is achieved by subtracting the mean value of each observation

and dividing by the standard deviation, as shown in Equation (1):

$$z = \frac{x - \mu}{\sigma} \dots\dots (1)$$

Where z is the transformed value of the feature, x is the original value of each descriptor, μ is the mean, and σ is the standard deviation of the feature in the dataset.

3.4. Data balancing using SMOTE

Data Balancing is a data pre-processing technique that handles uneven class distribution in a data set, where some classes have much more instances as compared to the other classes. Unbalanced data can introduce bias in model predictions and diminish classification performance. To overcome this issue, the SMOTE (Synthetic Minority Over-sampling Technique) method was applied to generate synthetic samples for the minority class. This process resulted in a balanced class distribution, thereby improving the accuracy, robustness, and reliability of the machine learning model.

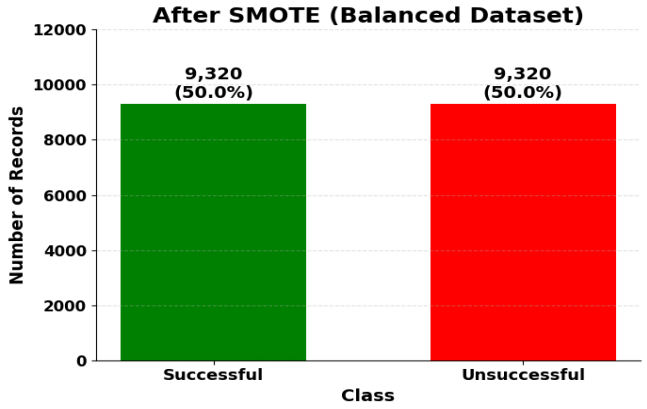

Fig 5: Bar Graph of Class Distribution after SMOTE

Fig. 5 The diagram illustrates the effect of applying the SMOTE technique to balance the dataset by generating synthetic samples for the minority class. As a result, the class distribution becomes more uniform, reducing prediction bias and improving the performance and reliability of the machine learning model.

3.5. Data Splitting

The dataset was split into training and testing subsets using stratified sampling with a 70:30 ratio. This approach preserved the original class distribution in both subsets, ensuring proportional representation and enhancing the reliability and generalization capability of the machine learning model.

3.6. Proposed Extreme Gradient Boosting (XGBoost) Model

In this work, a supervised machine learning-based boosting model, namely Extreme Gradient Boosting (XGBoost), is proposed for Clinical Trial Performance Prediction. XGBoost is an advanced ensemble learning algorithm that combines multiple decision trees in a boosting framework to improve predictive accuracy and reduce prediction errors. The model sequentially constructs decision trees, where each new tree learns from the residual errors of the previous trees, thereby enhancing overall performance. Owing to its built-in regularization mechanisms and efficient handling of complex data patterns, XGBoost is well suited for clinical trial outcome prediction. The mathematical formulation of the XGBoost model can be expressed as (2):

$$y = f(x) \dots\dots (2)$$

Where y is the predicted property price, x is the vector of input features (such as square footage, number of bedrooms, etc.), and f(x) is the XGBoost model that predicts y based on x. To compute f(x), XGBoost builds an ensemble of decision trees that are trained to minimize the mean squared error (MSE) loss function. The model combines the predictions from multiple decision trees to arrive at a final prediction. The general form of the XGBoost regression model can be expressed as (3):

$$y = \sum (k = 1\ to\ K)\ \ fk(x) \dots\dots (3)$$

Where fk(x) is the prediction of the k-th decision tree, and K is the total number of decision trees in the ensemble. The prediction of each tree is a weighted sum of the leaf values of the tree, which are learned during training. The prediction of the XGBoost model for a given input x is obtained by summing the predictions of all the decision trees in the ensemble.

The proposed Extreme Gradient Boosting (XGBoost) model was tuned using specific hyperparameter values to achieve optimal performance. The learning rate was set to 0.1, the maximum depth of trees was fixed at 6, and the number of estimators (trees) was set to 200. In addition, subsample was configured at 0.8 and colsample_bytree at 0.8 to improve generalization and reduce overfitting.

3.7. Evaluation metrics

To assess the effectiveness of classifiers, we employed established metrics, specifically accuracy, Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC), precision, recall, and F1-score. These metrics are grounded in the values of true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN):

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{Fp} + \mathrm{TN} + \mathrm{FN}} \dots\dots (4)$$

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \dots\dots (5)$$

$$Recall = \frac{\mathrm{TP}}{TP + FN} \dots\dots (6)$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \dots\dots (7)$$

Accuracy measures the proportion of correctly classified instances among all predictions made by the model. The Receiver Operating Characteristic (ROC) curve evaluates classifier performance by illustrating the trade-off between recall and the false positive rate across different classification thresholds, while the Area Under the Curve (AUC) summarizes this performance into a single value ranging from 0 to 1, with higher values indicating better classification capability. Precision assesses the reliability of positive

predictions by measuring the proportion of correctly identified positive instances among all predicted positives. Recall, also known as sensitivity, quantifies the model's ability to correctly identify actual positive instances, highlighting the impact of false negatives. The F1-score combines precision and recall into a single metric by calculating their harmonic mean, providing a balanced measure of classification performance, particularly when class distributions are uneven.

4. Results and Discussion

This section presents the experimental setup and evaluates the performance of the proposed model, demonstrating its effectiveness and efficiency in clinical trial performance prediction.

4.1. Experimental Setup

Experiments were conducted on a high-performance system, with an Intel Core i9-13900K (3.0 GHz) CPU, 64 GB of DDR5 RAM, and an NVIDIA RTX 4090 (24 GB VRAM) graphics card and is running Windows 11 Pro. Python libraries such as Pandas, NumPy, Matplotlib, Seaborn and Scikit-learn were used to implement and analyze the data.

4.2. Model Performance Evaluation

The proposed XGBoost model was trained and evaluated for clinical trial performance prediction in medical industry applications. Its performance was assessed using standard classification metrics, including accuracy, precision, recall, and F1-score, as presented in Table II. The model achieved an accuracy of 92.7%, demonstrating its ability to correctly classify the majority of instances. Additionally, it attained a precision of 95.6% and a recall of 95.9%, indicating high reliability in positive predictions and strong capability in identifying actual positive cases. The F1-score of 95.7% further reflects an excellent balance between precision and recall. These results demonstrate the effectiveness and robustness of the proposed XGBoost model for accurately predicting clinical trial performance outcomes in the medical industry.

Table 2: Classifiction Results of Proposed Model for Clinical Trial Performance Prediction

Matrix	Extreme Gradient Boosting (XGBoost)
Accuracy	92.7
Precision	95.6
Recall	95.9
F1-score	95.7

Fig. 6 presents the confusion matrix illustrates the classification performance of the proposed XGBoost model on the test dataset. The model successfully predicted 304 positive samples (True Positives) and 40 negative samples (True Negatives), showing good predictive performance. Moreover, it had a low misclassification rate with only 13 False Positives and 14 False Negatives. The accuracy of the XGBoost model was evidenced by the relatively small number of misclassified samples relative to the large number of samples classified correctly, indicating that the model proved to be an effective discriminator between positive and negative classes for achieving reliable and accurate prediction performance.

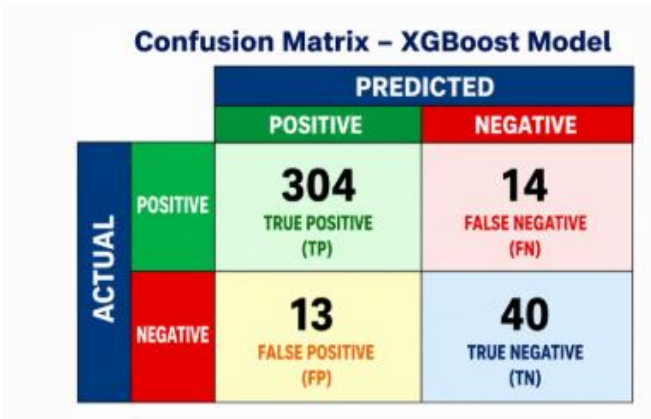

Fig 6: Confusion Matrix for the XGBoost Model

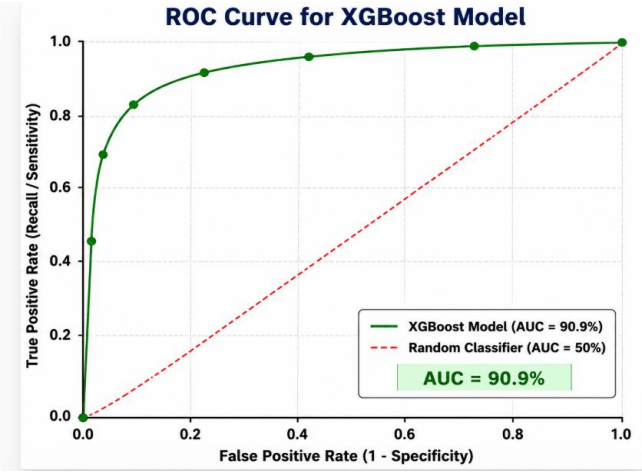

Fig 7: ROC Curve for XGBoost Model

Figure 7 shows the classification performance of the proposed XGBoost model using the ROC curve. The True Positive Rate (TPR; also known as the Sensitivity) is plotted against the False Positive Rate (FPR) at different decision thresholds. The curve is still close to the upper left corner of the graph, which shows that the classifier has good discrimination between positive and negative classes. The model achieved an Area Under the Curve (AUC) of 90.9%, demonstrating excellent predictive performance and a high ability to correctly distinguish between classes. The XGBoost model demonstrates strong classification performance, markedly outperforming the random classifier baseline, underscoring its robustness and reliability in the prediction task.

4.1. Comparative analysis

To evaluate the effectiveness of the proposed approach, a comparative performance analysis with several existing machine learning models was conducted, as shown in Table III. The results indicate that the proposed XGBoost model outperformed Logistic Regression (LR), Random Forest (RF), and CNN, achieving the highest accuracy of 92.7%, along with superior precision (95.6%), recall (95.9%), and F1-score (95.7%). Although the CNN model achieved a comparable

accuracy of 92.0%, its overall classification performance was lower than that of XGBoost. In contrast, RF and LR exhibited significantly lower predictive capabilities. These findings demonstrate the robustness and effectiveness of the proposed XGBoost model for clinical trial performance prediction in medical industry applications.

Table 3: Comparison of Different Machine Learning Models for Clinical Trial Performance Prediction

Model	Accuracy	Precision	Recall	F1-score
LR[28]	57.9	86.2	69.1	65.7
RF[29]	73.7	68.7	62.8	65.4
CNN[30]	92	89	94	-
XGBoost	92.7	95.6	95.9	95.7

4.2. Advantage of the Proposed Model and Discussion

The proposed XGBoost model demonstrated superior performance in clinical trial performance prediction, achieving the highest accuracy of 92.7% compared to the other evaluated machine learning models. Its key advantage lies in its ability to effectively capture complex relationships within clinical trial data while minimizing prediction errors through gradient boosting. The model also achieved high precision, recall, and F1-score values, indicating reliable identification of successful and unsuccessful trial outcomes with minimal misclassification. Compared with Logistic Regression, Random Forest, and CNN, the XGBoost model provided more robust and consistent predictive results, making it a suitable and efficient solution for supporting data-driven decision-making in medical industry applications.

5. Conclusion and future study

In order to explore the patient's intraoperative and postoperative adverse reactions, postoperative vital signs, and patient satisfaction in clinical trials of new sedative drugs, this study is based on experimental results. The comparative analysis of different machine learning models for clinical trial performance prediction shows that traditional Logistic Regression (LR) achieves the lowest accuracy of 57.9%, indicating limited capability in capturing complex data patterns. Random Forest (RF) improves performance significantly with 73.7% accuracy, while Convolutional Neural Networks (CNN) further enhance predictive ability, achieving 92% accuracy. The proposed XGBoost model outperforms all other approaches with the highest accuracy of 92.7%, demonstrating its superior ability to handle structured clinical trial data effectively. Overall, the results confirm that ensemble and deep learning methods provide better predictive performance, with XGBoost emerging as the most reliable model for clinical trial outcome prediction.

5.1. Limitations and Future Work

The study proposed is limited in that it relies upon one clinical trial dataset, potentially limiting the ability to extend this model to other medical domains and datasets. Further, the synthetic minority class data may lead to synthetic bias in the representation of the minority class in SMOTE. The study also has a narrow scope of only including structured clinical trial data, and neglecting unstructured data like clinical notes or images. Future work will focus on integrating multi-modal healthcare data and applying advanced deep learning and hybrid models to further improve prediction accuracy and robustness in real-world clinical trial environments.

References

[1] D. A. Berry *et al.*, "A cost/benefit analysis of clinical trial designs for COVID-19 vaccine candidates," *PLoS One*, vol. 15, no. 12, p. e0244418, Dec. 2020, doi: 10.1371/journal.pone.0244418.

[2] D. F. Heitjan, Z. Ge, and G. Ying, "Real-time prediction of clinical trial enrollment and event counts: A review," *Contemp. Clin. Trials*, vol. 45, pp. 26–33, Nov. 2015, doi: 10.1016/j.cct.2015.07.010.

[3] A. Warrier, "Real-Time Healthcare Event Processing: Stream Analytics for Clinical Decision Support," *Int. J. Emerg. Res. Eng. Technol.*, vol. 1, no. 4, Decmber, pp. 47–54, 2020, doi: https://doi.org/10.63282/3050-922X.IJERET-V1I4P106.

[4] X. Zhang and Q. Long, "Modeling and prediction of subject accrual and event times in clinical trials: a systematic review," *Clin. Trials*, vol. 9, no. 6, pp. 681–688, Dec. 2012, doi: 10.1177/1740774512447996.

[5] P. Kumar, "Leveraging Generative AI for Automated Data Standardization and Interoperability in Healthcare," in *2025 4th International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, Salem, India: IEEE, 2025, pp. 99–104, December. doi: 10.1109/ICAAIC64647.2025.11330217.

[6] H. N. Dholariya, "AI-Governed Data Modernization Architectures: A Secure and Compliant Framework for Healthcare and Life Sciences Cloud Ecosystems," *Front. Heal. Informatics*, vol. 15, no. 1, pp. 102–117, April, 2026, doi: https://doi.org/10.63682/fhi2984.

[7] R. Grout *et al.*, "Predicting disease onset from electronic health records for population health management: a scalable and explainable Deep Learning approach," *Front. Artif. Intell.*, vol. 6, Jan. 2024, doi: 10.3389/frai.2023.1287541.

[8] A. Rajkomar *et al.*, "Scalable and accurate deep learning with electronic health records," *npj Digit. Med.*, vol. 1, no. 1, p. 18, May 2018, doi: 10.1038/s41746-018-0029-1.

[9] B. Shickel, P. J. Tighe, A. Bihorac, and P. Rashidi, "Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis," *IEEE J. Biomed. Heal. informatics*, vol. 22, no. 5, pp. 1589–1604, 2017.

[10] A. Warrier, "Hybrid Cloud iPaaS for Healthcare Digital Transformation: Bridging On-Premises and Cloud-Based Health Information Systems," *Int. Sci. J. Eng. Manag.*, vol. 02, no. 01, pp. 1–9, Jan. 2023, doi: 10.55041/ISJEM00123.

[11] M. R. Anand and A. K. S, "Temporal Fusion Transformer Forecasting and MILP Prescriptive Optimization for Hospital Pharmacy Supply Chain Orchestration," in *2025 9th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, IEEE, Nov. 2025, pp. 1206–1213, Nov. doi: 10.1109/ICECA66444.2025.11382695.

[12] S. Xie, Z. Yu, and Z. Lv, "Multi-Disease Prediction Based on Deep Learning: A Survey," *Comput. Model.*

Eng. Sci., vol. 128, no. 2, pp. 489–522, 2021, doi: 10.32604/cmes.2021.016728.

[13] S. Mahmud, D. P. Mishra, G. . G. Ramani, M. I. Patel, M. S. Soumik, and R. Manivannan, "Design of intelligent healthcare IT infrastructure using graph theory, network analysis, and artificial intelligence," *Int. J. Appl. Math.*, vol. 38, no. 12s, pp. 2267–2280, december, 2025.

[14] M. van Smeden, J. B. Reitsma, R. D. Riley, G. S. Collins, and K. G. Moons, "Clinical prediction models: diagnosis versus prognosis," *J. Clin. Epidemiol.*, vol. 132, pp. 142–145, Apr. 2021, doi: 10.1016/j.jclinepi.2021.01.009.

[15] I. M. Putri, "ASUHAN KEPERAWATAN PADA TN.SI DENGAN CHRONIC OBSTRUCTIVE PULMONARY DISEASE (COPD) DI RUANG RAWAT INAP A RSUD KANJURUAN KEPANJEN," *Undergrad. thesis, Univ. Muhammadiyah Malang.*, 2023.

[16] X. Chen *et al.*, "Recent advances and clinical applications of deep learning in medical image analysis," *Med. Image Anal.*, vol. 79, p. 102444, Jul. 2022, doi: 10.1016/j.media.2022.102444.

[17] P. Kumar, "Edge Computing and IoT for Real-Time Healthcare Data Processing and Integration," in *2025 4th International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, Salem, India: IEEE, 2025, pp. 105–110, December. doi: 10.1109/ICAAIC64647.2025.11331211.

[18] M. Indirani, S. Sudheer, R. Mahaveerakannan, and P. Ruba, "Gallstone Disease Prediction Using Clinical and Biochemical Features Through Ensemble Learning Techniques," *Int. J. Comput. Intell. Syst.*, vol. 19, no. 1, p. 19, Dec. 2025, doi: 10.1007/s44196-025-01083-0.

[19] L. Vasudevan *et al.*, "Machine Learning Models to Predict Risk of Maternal Morbidity and Mortality From Electronic Medical Record Data: Scoping Review," *J. Med. Internet Res.*, vol. 27, pp. e68225–e68225, Aug. 2025, doi: 10.2196/68225.

[20] T.-T. Chen, "Predicting analysis times in randomized clinical trials with cancer immunotherapy," *BMC Med. Res. Methodol.*, vol. 16, no. 1, p. 12, Dec. 2016, doi: 10.1186/s12874-016-0117-3.

[21] R. Snehamrutha, "Patient Engagement Strategies in Community Pharmacies and their Effect on Vaccination Uptake and Medication Synchronizations," *ESP J. Eng. Technol. Adv.*, vol. 3, no. 3, pp. 163–173, September, 2023, doi: 10.56472/25832646/JETA-V3I7P120.

[22] T. Upadhaya, I. J. Chetty, B. Acharya, E. M. McKenzie, H. Bagher-Ebadian, and K. M. Atkins, "Machine Learning Models Predicting Radiation Pneumonitis Based on a Multi-regional Radiomics and Dosiomics Approach in Patients With Lung Cancer Treated on the NRG/RTOG 0617 Clinical Trial," *IEEE Trans. Radiat. Plasma Med. Sci.*, pp. 1–1, 2026, doi: 10.1109/TRPMS.2026.3663985.

[23] R. Baihaqi and I. Kurniawan, "Predictive Modelling of Clinical Trial Toxicity by Using Cuckoo Search-Ensemble Method," in *2025 International Conference on Information and Communication Technology (ICoICT)*, IEEE, Jul. 2025, pp. 1–6. doi: 10.1109/ICoICT66265.2025.11192977.

[24] U. S. Tasnim, S. Hossain, and M. M. Hasan, "Ensemble Machine Learning Models for Treatment Response Prediction and Adaptive Patient Allocation in Cancer Clinical Trials," in *2025 28th International Conference on Computer and Information Technology (ICCIT)*, IEEE, Dec. 2025, pp. 2034–2039. doi: 10.1109/ICCIT68739.2025.11490113.

[25] A. Devi *et al.*, "MIRLR: An ensemble approach for predicting clinical trial enrollment rates," in *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, IEEE, Jun. 2024, pp. 1–7. doi: 10.1109/ICCCNT61001.2024.10725806.

[26] M. Reinisch, J. He, C. Liao, S. Siddiqui, and B. Xiao, "CTP-LLM: Clinical Trial Phase Transition Prediction Using Large Language Models," in *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, Dec. 2024, pp. 3667–3672. doi: 10.1109/BIBM62325.2024.10822746.

[27] Y. Lu, S. Aslani, M. Emberton, D. C. Alexander, and J. Jacob, "Deep Learning-Based Long Term Mortality Prediction in the National Lung Screening Trial," *IEEE Access*, vol. 10, pp. 34369–34378, 2022, doi: 10.1109/ACCESS.2022.3161954.

[28] A. Iyer and S. Narayanaswami, "A Novel Model Using ML Techniques for Clinical Trial Design and Expedited Patient Onboarding Process," *Clin. Outcomes Res.* , vol. 17, no. January, pp. 1–18, 2025, doi: 10.2147/CEOR.S479603.

[29] B. Long, S.-W. Lai, J. Wu, and S. Bellur, "Predicting Phase 1 Lymphoma Clinical Trial Durations Using Machine Learning: An In-Depth Analysis and Broad Application Insights," *Clin. Pract.*, vol. 14, no. 1, pp. 69–88, Dec. 2023, doi: 10.3390/clinpract14010007.

[30] Sydney Anuyah, Mallika K Singh, and Hope Nyavor, "Advancing clinical trial outcomes using deep learning and predictive modelling: bridging precision medicine and patient-centered care," *World J. Adv. Res. Rev.*, vol. 24, no. 3, pp. 001–025, 2024, doi: 10.30574/wjarr.2024.24.3.3671.