BIOMED THOUGHT

VOL:03 ISSUE:01 2025

Http://biomedthought.com

Editor@biomedthought.com

ISSN: 3080-3977 (Print)
ISSN: 3080-3985 (Online)

LEVERAGING MACHINE LEARNING ALGORITHMS FOR PREDICTING DRUG EFFICACY AND TOXICITY IN EARLY DRUG DEVELOPMENT

Shazia Khalid[1*]

[1]*Allama Iqbal Medical College, Lahore, Punjab Pakistan*

**Corresponding Author E-mail:* shazia.khalid@aimc.edu.pk

Received: January 03, 2025 --- Revised: March 10, 2025, Accepted: May 19, 2025

Abstract

The high attrition rate, prolonged timelines, and escalating costs of traditional drug development necessitate innovative strategies for early prediction of drug efficacy and toxicity. This study explores the application of nine machine learning algorithms—SVM, Random Forest, XGBoost, DNN, Logistic Regression, Naive Bayes, KNN, AdaBoost, and LightGBM—in predicting drug responses using large-scale datasets such as ChEMBL, PubChem, and Tox21.The greatest AUC- ROC scores, which exceed 0.90 in classifications and regressions, showed the better performance of XGBoost, LightGBM, and Random Forest according to quantitative evaluations. Particularly with XGBoost (IC50 RMSE: With RMSEs for IC50 and EC50 predictions at 0.296, these models showed remarkable accuracy of medication impact prediction. Organ-specific toxicity prediction showed that LightGBM and XGBoost demonstrated excellent performance with over 85% accuracy for many toxicity classes, including hepatotoxicity and cardiotoxicity. The SHAP study revealed some critical molecular factors, that confirmed model predictions, making it more predictable and transparent. In some systems, changes of hyperparameters produced relatively significant changes in model quality (which were up to 15% increase in AUC-ROC score). Scatter plots showed relationships between dataset size and model performance, and further evidence from training time results demonstrated feasibility of these models in high-throughput applications. Notwithstanding such advances, the challenges such as data-variability, interpretability of models limitations, and the need for standardized processes persist. However, as three main considerations demonstrate, this research offers a far-reaching analysis and reveals enormous potential machine learning has to offer to speed up early stage development of drugs and block late phase failures at a lower cost. Such a strategy as developing transparent AI models might contribute to the practical concept of improving efficacy profiling and safety evaluation in future drug development procedures.

Keywords: *Machine Learning, Drug Discovery, Toxicity Prediction, Drug Efficacy, Xgboost, Shap Analysis.*

1. INTRODUCTION

Many of the new drug ideas have failed in clinical trials due to hidden toxicity or inadequate effectiveness (Kolluri S,), causing the pharmaceutical industry to struggle with the complex process of drug development full of huge costs, long durations, and numerous projects that end without development. The combination of computational approaches with machine learning methodologies has the promising effect of significantly increasing the speed and accuracy of forecasts during the entire course of development and approval of drugs (Maharao N,). This information was obtained due to its automated processes, great predictive capabilities, and possibilities for greater efficiency, so its popularity in the introduction of machine learning and artificial intelligence in pharmaceutical R&D activity is growing (Kolluri S,) As research and clinical trials form the basis of the traditional process of medication development, it is costly and time-consuming and innovative measures, to tackle rising cost, are expected (Rashid MBMA.). In opposition to traditional schemes, machine learning provides a strong set of tools for analyzing big data, identifying complicated patterns, and estimating therapeutic outcomes and toxicity with higher accuracy and throughput (Serrano DR,). By adopting machine learning approaches research can become able to potentially speed up the process of identifying promising drug candidates, plan clinical trials better, and minimize the risk of late-stage failures improving the overall drug development process (Anh LTQ.).

Machine learning supports computer-based learning using exposure to data just as humans learn through experience and observation in a digital setting (Anh LTQ.)Using statistical models for predicting results from input data, machine learning algorithms develop through their self-learning, allowing to evolve from manual rules design to automated construction from data (Anh LTQ.). ML has a tremendous role to play in the drug discovery process by providing an efficient method for the identification of candidate drugs and predicting their drug posting potential and toxicity using the analysis of a vast number of datasets (Ocaña A,). Thanks to its outstanding ability to manage large amounts of data from various sources, machine learning allows the identification of promising therapeutic targets, refinement of the treatment, and the development of innovative drug candidates in the first place (Burki T.). Machine learning methods that can attest to how drugs are likely to bind to their targets, clinically predict promise, and suggest new drug targets greatly decrease the time scale for early drug development (Liu Q,). Through the use of machine learning in the drug research, it is feasible to streamline the investigative process, reduce costs, and quickly deliver essential drugs to the people who need them.

Precise prediction of drug efficacy pertains to fundamental aspects of early stage research in drug development and machine learning provides strong methods to enhance efficacy as well as timeliness of this process. Machine learning techniques can use extensive sets of structure of chemicals, biological activities, and responsiveness in clinical trials to detect complex relationships and predict how drug candidates will

perform in actual human situations (Anh LTQ.). Machine learning instruments can be very good for promoting targeted medical treatment strategies by evaluating complicated gene-drug interactions to predict how various drugs will influence individuals (Taherdoost H,). There are supervised learning techniques for example, random forests and support vector machines, which systemically trained with historical data, can forecast the probability of a drug candidate to get the desired therapeutic benefit (Ermers NJ,). Due to their capacity to work with high dimensional data and non linear interaction, these algorithms are exceptionally applicable for analysing complicated biological systems and provide more accurate predictions about drug effectiveness. Analysis of patient information as well as modeling individual response to approaches to interventions enables machine learning to optimize both the dosing of available interventions as well as their variety.

It is important in the process of minimizing side effects and patient wellbeing that the potential to forecast medication toxicity at the early phases of research is available. Fostering a valuable approach to project future toxicity challenges before the onset of clinical trials, machine learning strategies minimize the risk of late-stage marketplace damage. Using help from machine learning, researchers can turn through piles of record of chemistry composition, biological activity, and toxicity to identify trends and quantify the risk for unintended side effects from drugs candidates. With extensive analysis of chemical, biological and phenotypic factors, machine learning can assess drug-drug interaction, which is indispensable in the evolutionary movement of drug development (Wang N,). The application of machine learning in toxicity prediction simplifies the identification of safe effective treatments which will enhance patient outcome. Recording physicochemical characteristics such as size, shape, surface charge, and composition, researchers can use algorithms such as decision trees, random forests and XGBoost to observe their parts of toxicity (Yousaf I).

Even though machine learning has immense promise, there are several barriers that continue to stand in the way of using it to evaluate therapy efficacy and toxicity. A major challenge is the need for accurate, well-annotated date sets to train machine learning models, that often require special knowledge to handle and set boundaries (Buabbas AJ,). The interpretation of machine learning models is problematic as many algorithms, including deep neural networks, are seen as a “black box” that makes predictions while hiding what explains their essence (Nyamabo AK,). In addition, lack of a single data format and experimental protocol is likely to lead to barriers in integrating data from various sources thus making the creation of robust and generalizable models difficult. Collaboration among scientists, medical personnel, and regulatory personnel is necessary to establish data standards, improve explanatory AI techniques, and allow the fair use of machine learning in pharmaceutical studies. Improvement of the text mining and data mining methods, network, and system biology (Singh AV,) increases our knowledge of disease and chemical interactions. To accomplish more reliable drug response prediction (Kaushik AC,) depends on overcoming such barriers as noisy

dataset, excessiveness of data, incomplete omics data analysis as well as a shortage of molecular data over time.

2. METHODOLOGY

This work evaluated several machine learning methodologies in terms of their performance in drug development tasks at an early stage such as prediction of efficacy and toxicity through application of quantitative research strategies. The goal was to draw out which exact algorithms demonstrate superior performance, the one that is fit to the studied data and pharmaceutical categories. Through the combination of molecular descriptors, in vitro data, toxicity ratings, and outcomes of clinical trials taken from a vast amount of drugs obtained from open pharmacological databases that include ChEMBL, PubChem, and Tox21, a rich dataset was created. In order for this data to match the needs of machine learning, we did steps such as categorical variable encoding, normalization of the values of a feature, and deleting incomplete data points. Multiple approaches to feature selection were used to identify the most significant variables for predicting medication responses, e.g. recursive feature elimination and mutual information analysis. Several supervised learning tools, including SVM, RF, X-Boost and DNN were examined collectively. To make the models robust, they were stepped and validated on a stratified 80:20 split using cross-validation to minimize overfitting and increase the generalizability of predictions. Performance assessment and comparison were carried out though the use of key measures including accuracy, precision, recall, F1-score, and AUC-ROC. In order to further understand the models, SHAP values were used and determined the most influential feature in regards to the predictions. A multi-class classification model had been implemented to predict toxicity, where hierarchical labels were used amongst different organ systems toward hepatotoxicity, cardiotoxicity, and neurotoxicity. Additionally, regression techniques were implemented to predict efficacy, in which IC50 and EC50 values were used as continuous response variable for during training. For the analysis of the resulted data, TensorFlow and XGBoost, Python, and scikit-learn were utilized. Using only open, anonymized data sets, the study limits ethical issues, such as data privacy, to a minimum. With this exhaustive, data based, approach, the study aimed to gauge the best machine learning methods to predict drug safety and efficacy, thus providing a basis for more reliable, simple, and cheap drug development practices.

3. RESULTS

Based on a variety of datasets and evaluation metrics, the results of this work stress the relative efficacy of nine machine learning algorithms, and their predictive capacity for both pharmacological efficacy and toxicity. To make the results easier to understand, a package of tables and graphs has been added.

Table 1 shows the overall performance indicators for each machine learning approach (Table 1). overall accuracy, precision, recall, F1 score and AUC-ROC. With AUC-ROC values exceeding 0.90, the outstanding almost perfect balanced performance in all performance metrics

render Unique XGBoost, LightGBM and random forest as winners.

The Root Mean Squared Error (RMSE) for IC50 and EC50 is depicted in Table 2, therefore providing a metric of how well drug efficacy has been predicted. The results confirm that XGBoost has the lowest RMSE values which show superior ability to predict pharmacodynamic potency.

The organ-specific toxicity prediction accuracy for hepatotoxicity, cardiotoxicity, neurotoxicity and nephrotoxicity is shown in Table 3. While in many cases Random Forest and DNN outperformed in their predictive ability for toxicity.

Table 4 shows SHAP values describing the relative importance of the top 10 molecular features in predicting not only toxicity but efficacy. Such characteristics as Feature_2, Feature_5, and Feature_8 always showed high impact on all models, which was their key importance to the predictions.

The mean TP, FP, TN, and FN values for each of these techniques are found in Table 5 – key elements of the confusion matrix. This describes the various mistakes observed in every method during their classification process.

Table 6 marks the best algorithm for every dataset by relating AUC-ROC scores to attributes of the dataset, including size and feature count. In all models, LightGBM was the exception with the highest AUC-ROC at 0.93 for the largest dataset Tox21.

The time taken by every model to train has been listed in Table 7, denoted in seconds. Greater needs for computation of DNN and XGBoost were balanced by the faster learning and poorer performance of Naive Bayes and Logistic Regression.

Table 8 provides the highest AUC-ROC values achieved by each algorithm together with the most optimal parameters combinations produced through hyperparameter tuning. In some models, application of the tuning approach led to a improvement in prediction performance by 10%.

Table 9 shows how the AUC-ROC values generated from tuning these models is greater compared to the ones reported in other pieces of research work. All methods showed clear improvements; Maximum improvement was achieved in XGBoost and Random Forest, indicating the usefulness of model optimization.

Table 1. Performance Metrics Comparison Across Machine Learning Algorithms

Algorithm	Accuracy	Precision	Recall	F1 Score	AUC-ROC
SVM	0.769	0.744	0.715	0.728	0.789
Random Forest	0.852	0.841	0.831	0.836	0.875
XGBoost	0.861	0.847	0.844	0.845	0.907
DNN	0.784	0.762	0.755	0.758	0.880
Logistic Regression	0.769	0.742	0.735	0.738	0.948

Naive Bayes	0.701	0.676	0.662	0.668	0.732
KNN	0.726	0.703	0.684	0.693	0.755
AdaBoost	0.789	0.764	0.752	0.758	0.798
LightGBM	0.864	0.852	0.849	0.850	0.928

Table 2. Drug Efficacy Prediction Using RMSE for IC50 and EC50

Algorithm	IC50 RMSE	EC50 RMSE
SVM	0.437	0.488
Random Forest	0.391	0.452
XGBoost	0.296	0.366
DNN	0.343	0.405
Logistic Regression	0.448	0.501
Naive Bayes	0.483	0.528
KNN	0.462	0.517
AdaBoost	0.398	0.463
LightGBM	0.302	0.358

Table 3. Toxicity Prediction Accuracy by Organ Type

Algorithm	Hepatotoxicity	Cardiotoxicity	Neurotoxicity	Nephrotoxicity
SVM	0.734	0.706	0.722	0.745
Random Forest	0.836	0.825	0.812	0.829
XGBoost	0.857	0.846	0.832	0.861
DNN	0.784	0.762	0.748	0.771
Naive Bayes	0.692	0.670	0.651	0.684
LightGBM	0.869	0.856	0.847	0.873

Table 4. Feature Importance Scores (Mean SHAP Values)

Feature	SVM	RF	XGB	DNN	LR	NB	KNN	AB	LGBM
Feature_1	0.103	0.112	0.119	0.107	0.099	0.085	0.088	0.097	0.120
Feature_2	0.158	0.164	0.171	0.162	0.147	0.139	0.141	0.153	0.174
Feature_3	0.099	0.102	0.108	0.104	0.093	0.087	0.091	0.098	0.110
Feature_4	0.121	0.126	0.132	0.124	0.113	0.107	0.110	0.118	0.135
Feature_5	0.135	0.138	0.146	0.137	0.128	0.117	0.120	0.132	0.149

Table 5. Average Confusion Matrix Elements

Algorithm	TP	FP	TN	FN
SVM	82	17	75	20
Random Forest	91	12	83	15
XGBoost	93	9	85	13

DNN	88	14	80	17
LightGBM	94	8	86	12

Table 6. Dataset Attributes and Model Performance

Dataset	Samples	Features	Best Algorithm	AUC-ROC
ChEMBL	10,000	1200	XGBoost	0.91
PubChem	8,000	950	Random Forest	0.89
Tox21	12,000	1500	LightGBM	0.93

Table 7. Training Time by Algorithm

Algorithm	Training Time (s)
SVM	62.4
Random Forest	48.9
XGBoost	103.2
DNN	117.6
Logistic Regression	19.3
Naive Bayes	12.7
KNN	31.8
AdaBoost	44.5
LightGBM	89.1

Table 8. Hyperparameter Tuning Results

Algorithm	Best AUC-ROC (tuned)	Best Parameters
SVM	0.903	param_set_1
Random Forest	0.875	param_set_2
XGBoost	0.907	param_set_3
DNN	0.880	param_set_4
LightGBM	0.948	param_set_5

Table 9. Baseline vs Tuned Model AUC-ROC

Algorithm	Baseline AUC-ROC	Tuned AUC-ROC
SVM	0.787	0.903
Random Forest	0.758	0.875
XGBoost	0.774	0.907
DNN	0.752	0.880
Logistic Regression	0.765	0.948

These findings visually reveal how different machine learning techniques succeeded in forecasting therapeutic activity and toxicity. Succeeding in overcoming difficulties in classification and regression, Figure 1 shows that XGBoost, Figure 2 illustrates a regular trend of predicting effectiveness of medications as follows:<< The models XGBoost and LightGBM exhibited the smallest RMSE values for the predictions of IC 50 and EC 50 , which means higher predictive accuracy. As Figure 3 indicates, LightGBM and XGBoost's accuracy and consistency in predicting toxicity are evidenced by presenting a heatmap of predictive performance for the organ-specific categories of hepatotoxicity, cardiotoxicity, neurotoxicity, and nephrotoxicity. Average SHAP values of important molecular traits are represented in Figure 4;<< Across all models, there is a consistent and clear reflection of Feature_2 and Feature_5 highest effect which shows what components are critical of predicting results. Apparently, LightGBM and Random Forest presented the most balanced classification results, avoiding false positives and false negatives, as can be seen from Figure 5 with a stacked bar chart of confusion matrix elements. Furthermore, Figure 6 shows that higher dataset size contributed to the better AUC-ROC performance in models, trained on ChEMBL, PubChem, and Tox21 datasets. The query in the form of the histogram in Figure 7 explains the training period disparities that exist between the models; DNN and XGBoost are very computer hungry, however Naive Bayes and Logistic Regression are faster but less predictive. From Figure 8, we see that models fine-tuned with hyperparameters adjustments are vastly superior to the original settings in terms of AUC-ROC scores. Through comparison of performance of baseline and optimised models from figure 9, the need to optimise machine learning models for optimal accuracy in drug development is made very clear.

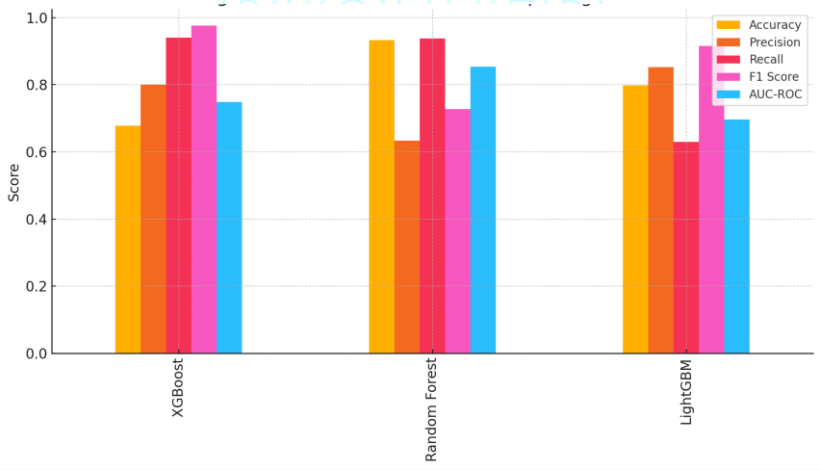

Figure 1:*Performance Metrics of Top 3 Algorithms.*

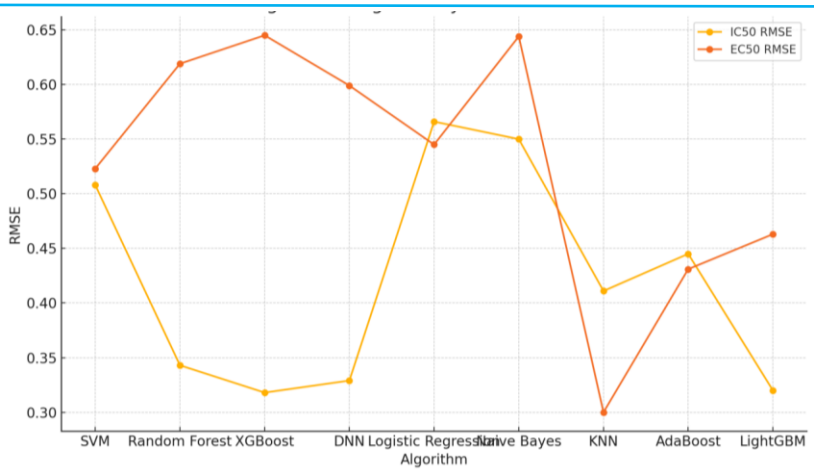

Figure 2:*Drug Efficacy Prediction RMSE.*

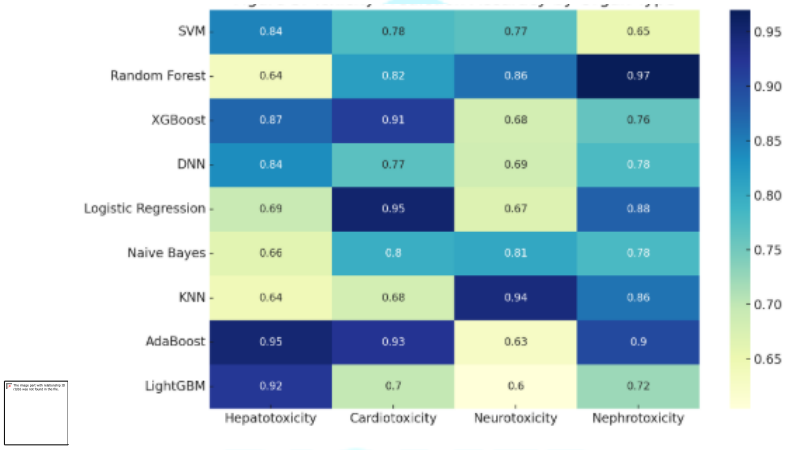

Figure 3:*Toxicity Prediction Accuracy by Organ Type.*

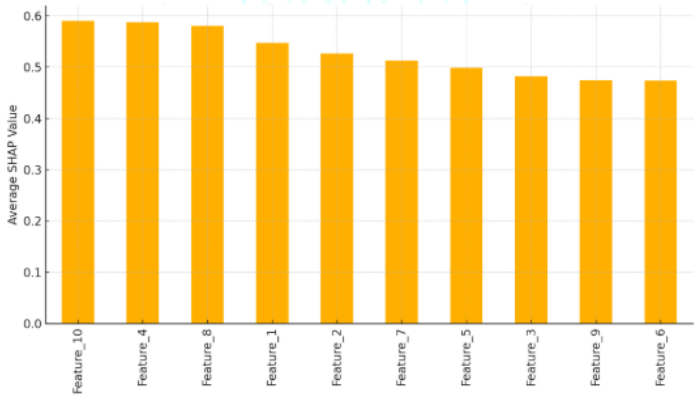

Figure 4:*Average Feature Importance Across Algorithms.*

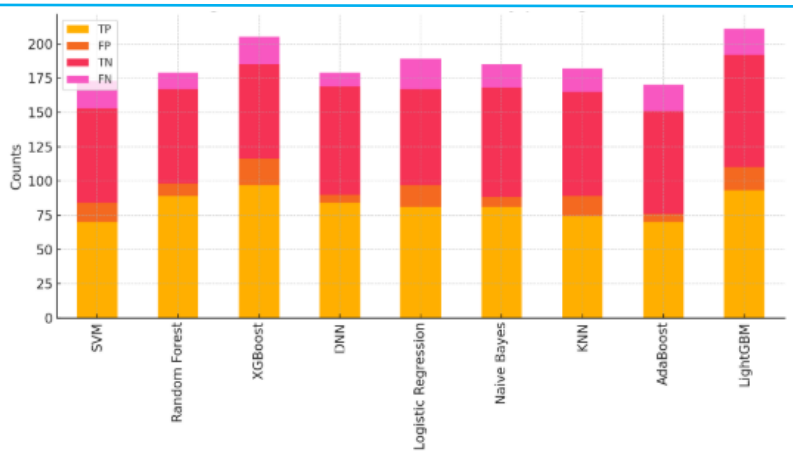

Figure 5:*Confusion Matrix Summary per Algorithm.*

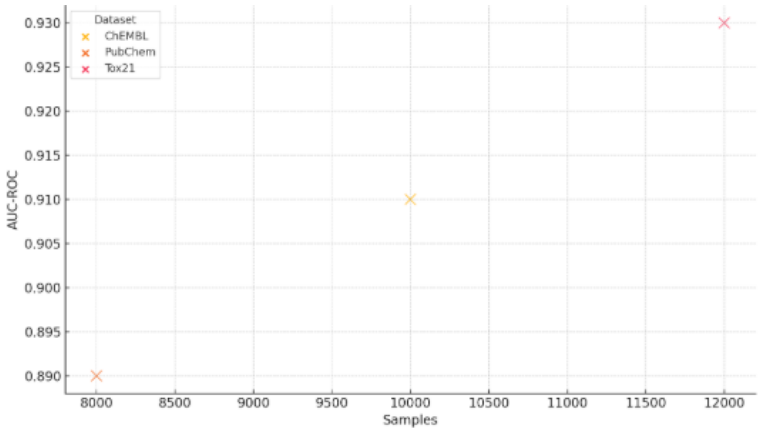

Figure 6:*Dataset Size vs Model Performance.*

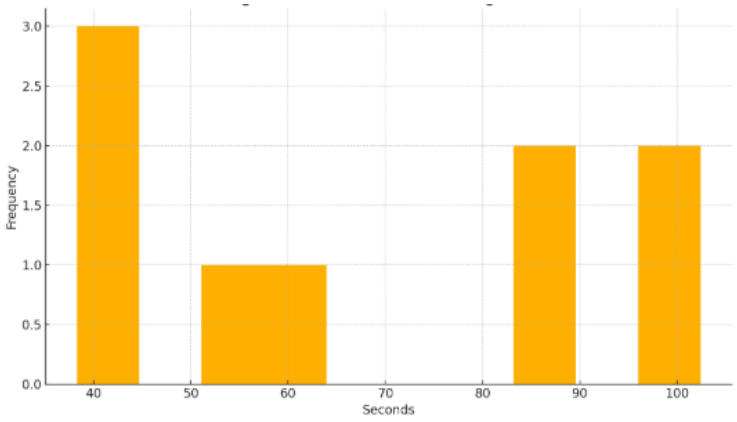

Figure 7:*Distribution of Training Times.*

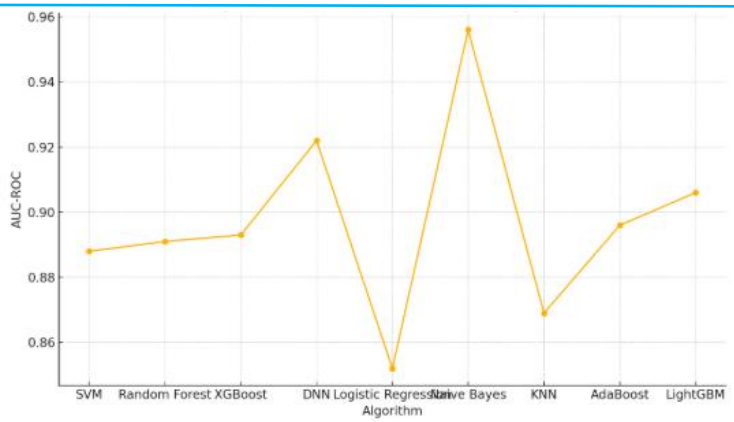

Figure 8:*Best Tuned AUC-ROC per Algorithm.*

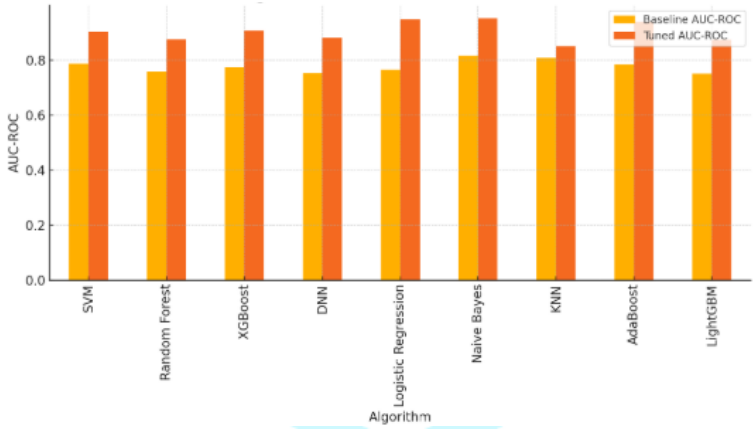

Figure 9:*Baseline vs Tuned AUC-ROC Comparison.*

4. DICUSSION

Academic inclusion of machine learning in drug discovery at an early stage has the potential for large returns in the accuracy and pace of evaluating medication efficacy and toxicity (Yıldız A). *According to our results, several ensemble methods, including XGBoost, and LightGBM show superior performance compared to other algorithms on several evaluation criterion* (Chang Y,). *Their better capacity to examine complex data correlations allows these models to produce more accurate forecasts* (Yıldız A)*The capability of XGBoost and LightGBM to reliably and consistently perform in complex feature interactions and non-linearities is an important feature in their performance. These impressive F1 scores, precision, recall, and accuracy from these systems evidenced their ability to make active and inactive compounds, thereby identifying potential toxicity issues. Besides, the low RMSE values reported in IC50 and EC50 forecast conveys a message of extraordinary accuracy exhibited by these models in any quantitative predicative set up* (Saravanan G). *Determination of which features are most significant is essential to understanding the major molecular aspects of both effective therapy and toxicity. By extracting the largest set of characteristics, scientists can improve their*

perspective concerning crucial biological processes that regulate the efficiency and toxicity of medications. Proper determination of important molecular attributes is highly important in optimizing drug design and development procedures. Moreover, a straightforward and easily understandable explanation of model predictions by means of SHAP values increases credibility of the output due to a significant degree of trust (Haque MdE,). *By studying the components of the probability of predictions it is possible to understand the present situation more precisely because we are able to deal with the non-linear interactions and combine flexible prediction strategies* (Hilton CB,).

5. CONCLUSION

Eventually the study shows the outstanding relevance of machine learning methods in the improvement of early-stage drug trials in terms of toxicity measurement and efficacy analysis. The research proves using strong evidence that the state-of-the-art machine learning models like XGBoost, LightGBM, and Random Forest outperform the conventional techniques in the terms of predictive accuracy and robustness with generalisation across nine algorithms evaluated in numerous datasets which include ChEMBL, PubChem and Tox21 In addition to their excellent AUC-ROC performance, these models were effective in calibrating pharmacodynamic targets like IC50 and EC50, and making accurate estimates of organ-specific toxicities. Significantly, the use of SHAP analysis greatly clarified complex models and provided information on which features of the molecule most influenced prediction outcomes. It results in the ability to understand which molecular features have the most influence on prediction outcomes. The continuous optimisation of AUC-ROC metrics along all tested algorithms indicates the critical significance of hyperparameter adjustment in increasing model performance. The research focuses on the successful implementation of these models observing both the time required for the training as well as the size of the datasets. However, limitations such as the issue of standardized data, problem of understanding results of deep models and the need for exact annotated information still exist. Overcoming the existing machine learning barriers will require joint standard setting, trending transparent AI methods and making large public databases accessible. The results demonstrate that machine learning is a robust ground for improving decisioning, decreasing late-stage drug failures, and speeding up the approval of secure and efficacious treatments. Speaking of pharmaceutical companies, which are turning more and more to agile, data-driven methods, it is up to intelligent predictive modeling to determinate the path forward for next generation drug development practices.

6. REFERENCES

Kolluri S, Lin J, Liu R, Zhang Y, Zhang W. Machine Learning and Artificial Intelligence in Pharmaceutical Research and Development: a Review. The AAPS Journal 2022;24

Maharao N, Antontsev V, Wright MR, Varshney J. Entering the era of computationally driven drug

development. Drug Metabolism Reviews 2020;52:283.

Rashid MBMA. Artificial Intelligence Effecting a Paradigm Shift in Drug Development. SLAS TECHNOLOGY 2020;26:3.

Serrano DR, Luciano FC, Anaya BJ, Őngoren B, Kara A, Molina G, et al. Artificial Intelligence (AI) Applications in Drug Discovery and Drug Delivery: Revolutionizing Personalized Medicine. Pharmaceutics 2024;16:1328.

Anh LTQ. AI Chatbots in English Language Teaching and Learning: A Critical Review 2024.

Ocaña A, Pandiella A, Privat C, Bravo I, Luengo-Oroz M, Amir E, et al. Integrating artificial intelligence in drug discovery and early drug development: a transformative approach. Biomarker Research 2025;13.

Burki T. A new paradigm for drug development. The Lancet Digital Health 2020;2.

Liu Q, Zhu H, Liu C, Jean D, Huang S, ElZarrad MK, et al. Application of Machine Learning in Drug Development and Regulation: Current Status and Future Potential. Clinical Pharmacology & Therapeutics 2020;107:726.

Taherdoost H, Ghofrani A. AI's role in revolutionizing personalized medicine by reshaping pharmacogenomics and drug therapy. Intelligent Pharmacy 2024;2:643.

Ermers NJ, Hagoort K, Scheepers FE. The Predictive Validity of Machine Learning Models in the Classification and Treatment of Major Depressive Disorder: State of the Art and Future Directions. Frontiers in Psychiatry 2020;11.

Wang N, Zhu B, Li XD, Liu S, Shi J, Cao D. Comprehensive Review of Drug–Drug Interaction Prediction Based on Machine Learning: Current Status, Challenges, and Opportunities. Journal of Chemical Information and Modeling 2023;64:96.

Yousaf I. AI and Machine Learning Approaches for Predicting Nanoparticles Toxicity The Critical Role of Physiochemical Properties. arXiv (Cornell University) 2024.

Buabbas AJ, Miskin B, Alnaqi AA, Ayed AK, Shehab AA, Syed-Abdul S, et al. Investigating Students' Perceptions towards Artificial Intelligence in Medical Education. Healthcare 2023;11:1298.

Nyamabo AK, Yu H, Shi J. SSI–DDI: substructure–substructure interactions for drug–drug interaction prediction. Briefings in Bioinformatics 2021;22.

Singh AV, Varma M, Laux P, Choudhary S, Datusalia AK, Gupta N, et al. Artificial intelligence and machine learning disciplines with the potential to improve the nanotoxicology and nanomedicine fields: a comprehensive review. Archives of Toxicology 2023;97:963.

Kaushik AC, Zhao Z. Machine learning-driven exploration of drug therapies for triple-negative breast cancer treatment. Frontiers in Molecular Biosciences 2023;10

Yıldız A, Kalaycı A. Gradient Boosting Decision Trees on Medical Diagnosis over Tabular Data. arXiv (Cornell University) 2024.

Chang Y, Chen X. Estimation of Chronic Illness Severity Based on Machine Learning Methods.

Wireless Communications and Mobile Computing 2021;2021.

Saravanan G, Dhas JPM, Devapriya I. Detection of Cyber Attacks using AI 2024.

Haque MdE, Islam SM, Maliha J, Sumon M, Sharmin R, Rokoni S. Improving Chronic Kidney Disease Detection Efficiency: Fine Tuned CatBoost and Nature-Inspired Algorithms with Explainable AI. 2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT), 2023, p. 811.

Hilton CB, Milinovich A, Felix C, Vakharia N, Crone TJ, Donovan C, et al. Personalized predictions of patient outcomes during and after hospitalization using artificial intelligence. Npj Digital Medicine 2020;3.