Personalized Drug Recommendation using Ensemble Machine Learning: A Predictive Analytics Approach

Rohan Kumar Rai
Dept.of Computer Engineering
Marwadi University
Rajkot, Gujarat
rohankumarrai2003@gmail.com

Ashish Kumar Singh
Dept.of Computer Engineering
Marwadi University
Rajkot, Gujarat
ashishkumar77393@gmail.com

Om Prakash Suthar*
Dept.of Computer Engineering
Marwadi University
Rajkot, Gujarat
opsuthar18@gmail.com

***Abstract*— Symptoms of asthma include wheezing, difficulty breathing, chest tightness, and coughing, which are caused by an inflammatory condition of the airways. Every day, it lowers the standard of living for millions of people around the globe. The severity of asthma varies among patients, necessitating personalized treatment strategies to ensure effective symptom management. Conventional therapy regimens frequently take a one-size-fits-all approach, which might not be the best for every patient. This study presents a Personalized Drug Recommendation System designed to assist in asthma treatment decisions by leveraging machine learning techniques. The system evaluates patient health data before and after administering Drug-S, identifying trends in key health indicators such as lung function, oxygen levels, and symptom severity. A dataset of asthma patients was analyzed using Exploratory Data Analysis (EDA), feature engineering, and model evaluation techniques. Three machine learning models—Random Forest (Accuracy:83%), XG-Boost (Accuracy:85%), and a Hybrid Model (Accuracy:85%)—were implemented and compared based on performance metrics such as accuracy, precision, recall, and F1-score. The results demonstrate the effectiveness of predictive modeling in optimizing asthma treatment by providing data-driven insights for better decision-making.**

***Index Terms*—Electronic Health Record, Classification, Feature Selection, Machine Learning, Extreme Gradient Boosting, Exploratory Data Analysis.**

I. INTRODUCTION

Asthma is a widespread chronic respiratory disease that affects millions of people worldwide, requiring continuous monitoring and personalized treatment strategies to ensure effective management. It is characterized by airway inflammation, which leads to breathing difficulties, wheezing, chest tightness, and persistent coughing. The condition significantly impacts daily activities, work productivity, and overall quality of life, placing a heavy burden on public health systems globally. [7] Managing asthma effectively demands timely intervention, proper medication, and lifestyle adjustments to minimize symptoms and prevent severe complications. Inadequate or improper treatment can result in frequent exacerbations, hospitalizations, and even life-threatening conditions such as acute asthma attacks. Therefore, selecting the right medication tailored to the patient's condition is critical for ensuring symptom control and long-term well-being [8], [11].

Governments and healthcare organizations worldwide recognize the need for accessible and affordable asthma treatments and provide essential medications through various public health programs. These medications typically include bronchodilators (such as beta-agonists), inhaled corticosteroids (ICS), leukotriene modifiers, and combination therapies designed to reduce inflammation and improve airflow. Physicians prescribe these treatments based on individual patient profiles, considering factors like asthma severity, age, medical history, and potential triggers. However, the one-size-fits-all approach to asthma treatment often proves inefficient, as patients respond differently to medications due to variations in genetics, environmental factors, and comorbidities. As a result, personalized medicine and data-driven drug recommendation systems are gaining importance in optimizing asthma treatment [9].

Leveraging machine learning, artificial intelligence, and real-world patient data could help develop tailored treatment plans, improving medication efficacy and reducing adverse effects. Further advancements in pharmacogenomics, digital health monitoring, and AI-driven healthcare solutions hold promise for enhancing asthma management strategies, ensuring that patients receive the most effective, personalized, and accessible treatments for their condition [10] [16].

A. Motivation

The situation of increasing complex and chronic diseases together with noticeable individual differences in drug effects requires new personalized healthcare solutions specifically for pharmacotherapy. Current drug prescription practices based on standardized clinical guidance and a single treatment approach generate less-than-optimal outcomes as well as drug side effects and delayed results and increased medical costs [12], [13]. In polypharmacy scenarios with comorbidities the danger of medication interactions as well as the use of improper medications becomes more significant. The drug recommendation system aligns drug choices with individual patient needs using specific information that includes DNA sequences along with existing health records and current wellness measurements and life patterns. Built into these platforms, machine learning algorithms and biological data analytics can change clinical decision-making and improve patient safety, hence providing tailored treatment alternatives [14].

B. Contribution

In this study, we look at advanced machine learning methods that can recommend personalized drug to patient on the basis of specific symptoms [15]. The article's key arguments are as follows:

- We offer a novel data-driven method that integrates

demographics, medical history, and genetic data with patient-specific traits to generate personalized medication recommendations with the goal of optimizing therapeutic outcomes.

- Through the application of state-of-the-art machine learning methods, such as ensemble models and DL architectures, the structure is able to process intricate biomedical data and determine the most effective and safe pharmacological solutions for specific patients.
- Our system combines EHRs with patient genomic data and drug interaction information within a diverse database for generating precise clinically oriented suggestions.
- The evaluation of the proposed model reveals superior correctness and precision in addition to recall enhancement when predicting suitable drug treatments for patient profiles.

C. Organization

A general outline of the paper's format is as follows. The current research and its conclusions are presented in Section 2. In Section 3, we offer a methodology framework and provide an overview of machine learning algorithms. In Section 4, we learn about the study and evaluation of performance with different parameters using a graph, a confusion matrix, and an accuracy table. Lastly, in the sections that follow, we wrap up our future works.

II. LITERATURE REVIEW/SURVEY

A drug recommendation system for multi-disease patients using machine learning is presented in the 2020 research article "A Drug Recommendation System for Multi-disease in Health Care Using Machine Learning" [6]. This study makes use of a dataset that was built for testing reasons and includes symptoms of heart troubles, common colds, fever, obesity, eye problems, and orthopedic problems. Support Vector Machine (SVM), Decision Tree, Random Forest, and K-Nearest Neighbors (KNN) are some of the supervised machine learning algorithms used to build the recommendation system. Random Forest outperformed the other models in terms of accuracy for drug recommendation, reaching 96.87%.

The study "Disease Prediction and Drug Recommendation using Machine Learning" (2024) [1] focuses on developing a predictive model for disease diagnosis and drug recommendations. The research utilizes datasets from Columbia University for disease prediction and Kaggle for disease-related symptoms, incorporating patient symptoms, medical history, demographic data, and drug reviews as key features. Multiple Supervised Machine Learning algorithms, including Logistic Regression, Random Forest, K-Nearest Neighbors (KNN), Naïve Bayes, Support Vector Machine (SVM), and Decision Tree, are applied to analyze patient data. The model effectively predicts diseases and suggests appropriate drug recommendations based on patient conditions.

The research paper "Drug Recommendation System Based on Sentiment Analysis of Drug Reviews Using Machine Learning" (2024) [2] explores the use of sentiment analysis for drug recommendations by analyzing patient reviews and ratings. The study utilizes the UCI Machine Learning Repository - Drug Review Dataset, which contains patient feedback, drug effectiveness ratings, and sentiment-based insights. To process the data, the study employs word embeddings and manual feature analysis to extract meaningful information from text-based reviews. Several Supervised Machine Learning algorithms are implemented, including Linear SVC, Naïve Bayes, SVM, Logistic Regression, Decision Tree, and SGD Classifier. Among these, Linear SVC with TF-IDF achieved the highest accuracy of 93%, whereas Decision Tree with Word2Vec had the lowest accuracy at 78%.

The study "Causal Inference-Driven Medication Recommendation with Enhanced Dual-Granularity Learning" (2024) [4] introduces a causality-based approach to medicine suggestion utilizing MIMIC-III and MIMIC-IV databases of electronic health records. The model considers patient history, medication effects, and molecular-level drug interactions to enhance personalized drug prescriptions. The research leverages Graph Neural Networks (GNNs) to model the relationships between diseases, medications, and their effects. This approach improves decision-making by incorporating both coarse-grained (general medication effects) and fine-grained (molecular-level drug interactions) learning techniques. The proposed model achieves a 2.54% accuracy improvement over state-of-the-art models, a 3.65% reduction in side effects (lower Drug-Drug Interaction rate), and a 39.42% increase in computational efficiency. Future work aims to enhance model scalability and integrate reinforcement learning techniques for better medication optimization.

The research paper "A Comprehensive Pharmaceutical Recommendation System" (2024) [3] presents a hybrid approach for pharmaceutical recommendations using data extracted from Drugs.com and Druglib.com. The study incorporates patient details (age, gender, medical history) and drug-related features (categories, interactions, side effects, and effectiveness ratings) to enhance drug recommendation accuracy. The model employs a Hybrid Recommendation System combined with Modified K-Means clustering to categorize drugs based on their effectiveness and patient suitability. The proposed approach significantly improves recommendation performance, achieving a 31% increase in accuracy, a 29% improvement in sensitivity, and a 28% rise in hit rate compared to conventional machine learning models. Below mentioned table I and II represents summarized information of various published articles on this research.

TABLE I LITERATURE REVIEW ON RELATED ARTICLES

YEAR	DATASET	FEATURES	METHOD	RESULT
2020 [6]	Sample dataset (test purposes)	Symptoms of multiple diseases (cardiac, common cold, fever, obesity, optical, ortho)	SVM, Random Forest, Decision Tree, KNN	Random Forest (96.87% accuracy)
2024 [1]	Columbia University (disease prediction), Kaggle (disease symptoms)	Symptoms, medical history, demographic data, drug reviews	Logistic Regression, Random Forest, KNN, Naïve Bayes, SVM, Decision Tree	Accurate disease prediction and drug recommendations
2024 [2]	UCI Machine Learning Repository - Drug Review Dataset	Reviews, ratings, sentiment analysis, word embeddings, manual feature analysis	Linear SVC, Naïve Bayes, SVM, Logistic Regression, Decision Tree, SGD Classifier	Linear SVC with TF-IDF (93% accuracy), Decision Tree (78% accuracy)
2024 [4]	MIMIC-III, MIMIC-IV (EHR)	Patient history, medication effects,	Graph Neural Networks (GNN)	2.54% improvement in accuracy, 3.65%

		molecular drug interactions		reduction in side effects, 39.42% efficiency improvement
2024 [3]	Drugs.com, Druglib.com	Patient details (age, gender, history), Drug details (categories, interactions, side effects)	Hybrid Recommendation System, Modified K-means	31% increase in accuracy, 29% sensitivity improvement, 28% hit rate increase
2022 [5]	European healthcare data	Privacy vs access trade-offs	GDPR	Privacy laws limit AI advancements

TABLE II COMPARISON ON RELATED ARTICLES

Cite	Sup-ML	Unsup - ML	DL	Sentiment Analysis	Graph -Based Model	Real-World Dataset	High Accuracy
[6]	YES	NO	NO	NO	NO	NO	YES
[1]	YES	NO	NO	NO	NO	YES	YES
[2]	YES	NO	NO	YES	NO	YES	YES
[4]	NO	NO	NO	NO	YES	YES	YES
[3]	NO	YES	NO	NO	NO	YES	YES
[5]	NO	NO	NO	NO	NO	YES	NO

III. PROPOSED METHODOLODY

In this part, we lay out the machine learning methods and explain how the suggested framework will be used to implement the medication recommendation system.

A. Dataset used:

The dataset comprises 18,215 patient records, each with 21 attributes, centered around asthma treatment and medication adherence. It includes various key features such as patient demographics like age and gender (female), along with important details from medical history including the use of previous asthma drugs, the number of pre-asthma days, and post-index exacerbations over a 365-day period. The dataset also captures comorbid conditions such as pneumonia, sinusitis, acute bronchitis, upper respiratory infection, and rhinitis, which may influence treatment decisions. Furthermore, it provides insights into medication usage and treatment costs, covering variables like the total number of pre-index inhaler canisters used in a year, adherence to prescribed medications, pre-asthma-related charges, pharmaceutical expenses, and the total cost incurred before the index period. The target variable in this dataset is 'drugs', representing the recommended type of drug for each patient based on the provided clinical and behavioral attributes. Table III shows various parameters and its measurements units.

TABLE III UNITS OF MEASUREMENT

PARAMETER	UNIT	DESCRIPTION
Lung Function	L/sec	Liters per second
Oxygen Level	%	Blood oxygen saturation
Age	Years	Patient's age
Drug s	Binary(0/1)	0= Not Given, 1= Given

B. Proposed Methodology Architecture:

The proposed methodology architecture illustrated in the flowchart outlines the structured machine learning pipeline used for developing a Personalized Drug Recommendation System for Asthma Patients. As mentioned in figure 1, the process begins with collecting patient details, followed by attribute selection and EDA (Exploratory Data Analysis) to recognize the dataset.

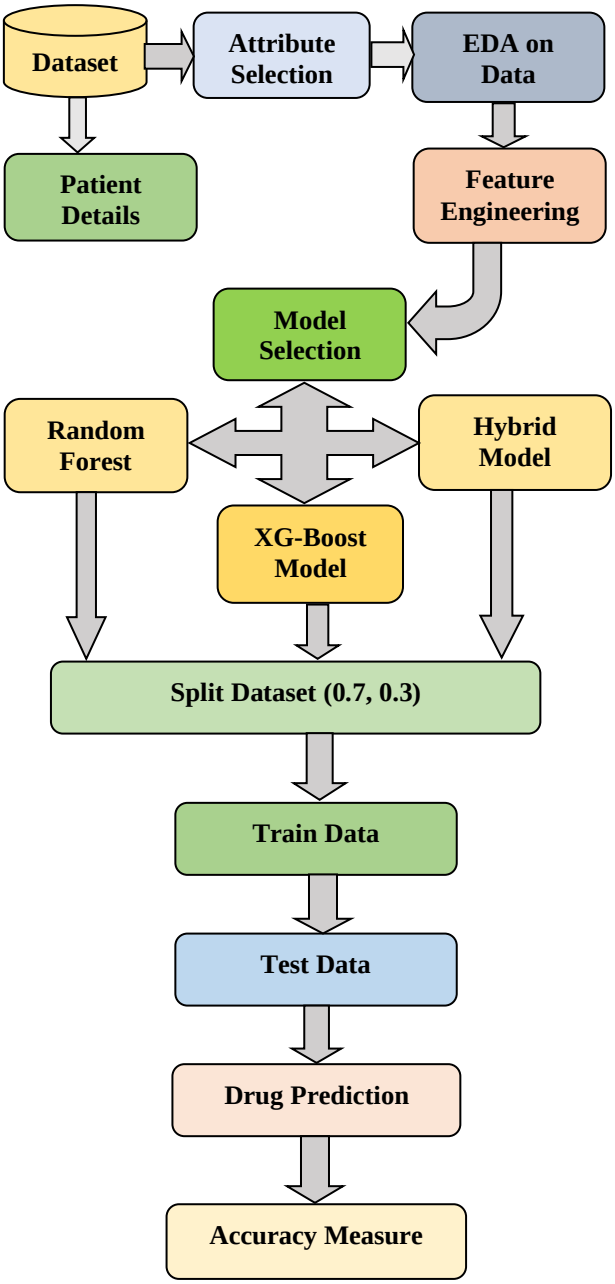

Fig. 1. Proposed Architecture of Drug Recommendation System

Subsequently, feature engineering is applied to transform and enhance data quality for model training. The system then proceeds to model selection, comparing three approaches—Random Forest, XG-Boost, and a Hybrid Model combining multiple algorithms. Partitioning the dataset into training (representing 70% of the total) and testing (30%) sections follows model selection. The model is constructed using the training set, and the test set evaluates its performance. The models then predict the most suitable asthma medication (Drug S) for each patient. Finally, the system assesses prediction effectiveness using accuracy measures, ensuring the model's reliability and practical applicability in personalized asthma treatment.

C. Exploratory Data Analysis (EDA) and Feature Engineering:

A process known as exploratory data analysis (EDA) is carried out before the training of the model in order to identify trends within the dataset. The process of converting and sanitizing data is carried out in two stages: The process of managing missing values is denoted by the first step: Imputation techniques such as mean, median, and KNN are utilized in order to fill in missing entries. Additional steps include the process of feature normalization, which involves the modification of numerical features through the use of min-max normalization [13].

$$X\_scaled = \frac{X\_max - X\_min}{X - X\_min} \quad (1)$$

Categorical Encoding: Categorical variables (e.g., gender) are converted into numerical format using one-hot encoding. Next, Feature Selection and Correlation Analysis consist of a correlation matrix that is used to determine dependencies between features. Features with high correlation (r > 0.75) are eliminated to avoid multicollinearity.

D. Model Implementation:

The study implements three machine learning models to predict the effectiveness of Drug-S.

Random Forest (RF) Model:

The construction of several decision trees by means of a model that employs ensemble learning strategies. Through the utilization of the *n_estimators'* option, the model is able to determine the number of decision trees that are utilized. Node splitting is accomplished by the utilization of the Gini Index, and Equation 2 provides an illustration of the computation method for the Gini Index.

$$GI = 1 - \sum_{i-1}^{n} (P_i)^2 \quad (2)$$

where P_i is the probability of class *i*.

XG-Boost (XGB) Model:

XGBoost is a method that builds an ensemble of decision trees sequentially, where each new tree focuses on correcting the errors made by the previous one. It employs sophisticated optimization strategies and regularization approaches that mitigate overfitting and enhance model efficacy. A gradient boosting model that minimizes the loss function using equation 3:

$$\mathcal{L} = \frac{1}{N} \sum_i \ell(\hat{y}_i, y_i) + \sum_j \Omega(f_j) \quad (3)$$

Hybrid Model (RF + XGB + SVM):

Integrates Random Forest, XG-Boost, and Support Vector Machine (SVM) to enhance generalization. Utilizing the SVM Kernel Trick: Employs Radial Basis Function (RBF) Kernel which is shown in equation 4:

$$K(X, Y) = exp(-\frac{||X - Y||^2}{2\sigma^2}) \quad (4)$$

IV. RESULTS AND DISCUSSION

This section shows the various performance evaluation matrices and results of various models with their analysis.

A. Performance Evaluation Metrices:

To assess model performance, key evaluation metrics are used:
Accuracy: Measures the proportion of correct predictions.
Precision and Recall: Precision measures how many predicted positive cases were actually positive while Recall measures how many actual positive cases were correctly identified.
F1-Score: Balances precision and recall.

$$Accuracy = \frac{tp + tn}{tp + tn + fp + fn} \quad (5)$$

$$Precision = \frac{tp}{tp + fp} \quad (6)$$

$$Recall = \frac{tp}{tp + fn} \quad (7)$$

$$F1\_Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (8)$$

To improve model accuracy and avoid errors, common pitfalls are addressed in table IV:

TABLE IV PROBLEMS DURING PERFORMANCE EVALUATION

COMMON MISTAKE	***SOLUTION***
Overfitting	Use **cross-validation**, ensemble learning
Imbalanced Data	Use **SMOTE (Synthetic Minority Over-sampling Technique)**
Multicollinearity	Remove **highly correlated features**
Incorrect Feature Scaling	Use **standardization or normalization**

B. Performance Analysis and Results

Our suggested approach uses three ML algorithms on the input dataset, which signifies drug recommendation on the basis of various key features with strong relationships by using correlation matrix. Figure 2 shows correlation matrix:

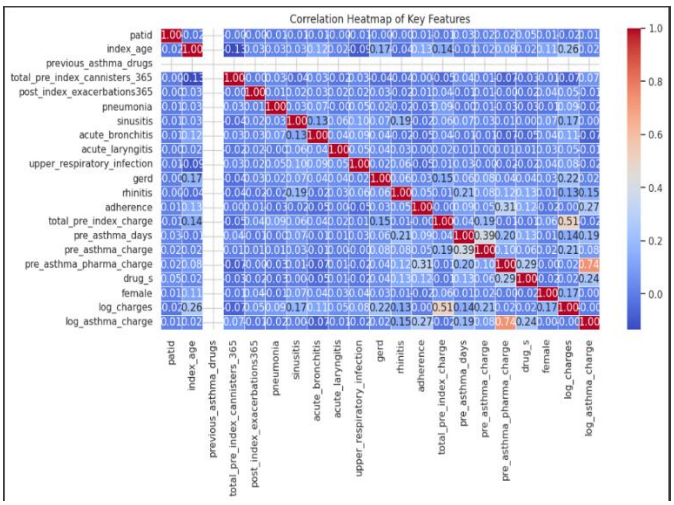

Fig. 2. Correlation Matrix of various key features

Below mentioned figure 3(a) is the age distribution chart which shows a bimodal pattern suggesting that the dataset mostly comprises two main age groups: young people about 10–15 years and middle-aged adults approximately 40–50 years. The histogram shows a clear drop in representation for the 20–30 age range, implying a possible underrepresentation of that group in the data. As models could become biased toward predicting outcomes more accurately for the age groups with larger representation, this unequal distribution could affect the performance of the personalized drug prescription system. A smooth KDE curve serves to show the general density tendency, hence stressing the age distribution variation and the need of guaranteeing balanced learning -

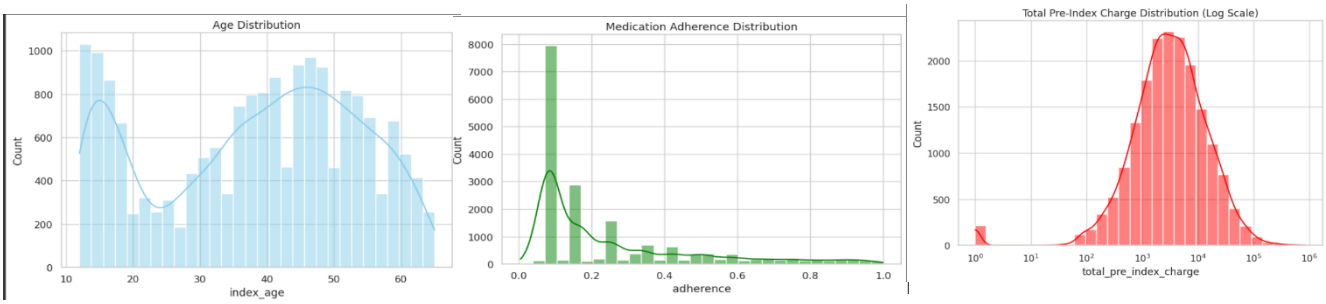

Fig. 3. key features in distribution (a) Age Distribution Graph, (b) The Medication Adherence Distribution, (c) Total Pre-Index Charge Distribution graph

across all age groups for consistent and fair drug recommendations.

The Medication Adherence Distribution graph shown in figure 3(b) which reveals a strongly right-skewed distribution suggesting most patients cluster around the 0.0–0.1 range with very low adherence levels. A sharp peak close to zero implies that many people seldom follow their prescribed drug regimen. With very few patients attaining adherence levels over 0.5, the number of people declines dramatically as compliance rises. Low adherence could obscure genuine treatment results; hence this distorted behavior could significantly affect drug efficacy and model projections. Improving the accuracy and fairness of a tailored medicine recommendation system depends on hence tackling adherence variation.

Figure 3(c) Plotted on a logarithmic scale, the Total Pre-Index Charge Distribution graph shows a near-normal (bell-shaped) distribution concentrated around the mid-range of charges, roughly between 10^3 and 10^4 (i.e., 1,000 to 10,000 units). This implies that most people had reasonable healthcare costs before treatment indexing; fewer suffered either very low or very high prices. A log scale aids in condensing the broad spectrum of numbers and exposing underlying patterns that might otherwise be concealed in a linear graphic. Patients' varied cost load shown by this distribution could represent different disease severity, treatment intensity, or access to healthcare. Knowing this charge distribution is crucial since it can greatly affect drug recommendation and tailored therapy plans.

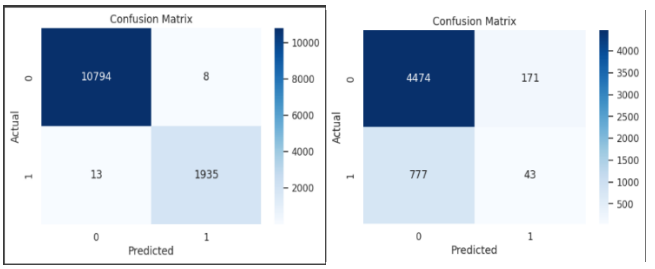

Fig. 4. (a) Confusion matrix of Random Forest Model (Train & Test datasets)

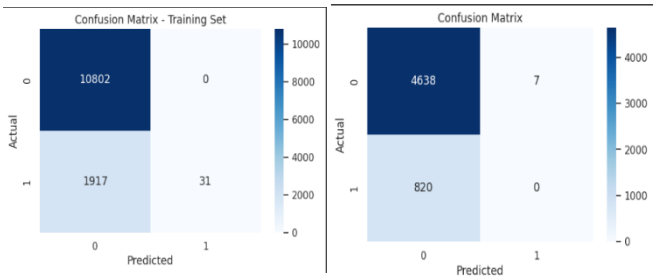

Fig. 4. (b) Confusion matrix of Hybrid Model (Train & Test datasets)

Figure 4(a) and 4(b) demonstrates the confusion matrix of Random Forest model and Hybrid model respectively. In testing evaluation of random forest model tp=4474, tn=43, fp=171 and fn=777 predicted which signifies accuracy as 82.65%, precision as 96.31%, recall as 85.20% and F1-score as 90.41%. Similarly, if we check confusion matrix of hybrid model for testing evaluation, we find the accuracy as 84.86%, precision as 99.84%, recall as 84.97% and F1-score as 91.80%.

The below mentioned Figure 5 demonstrates accuracy graph of various model, the personalized drug recommendation system accuracy assessment demonstrates important performance trends between Random Forest and XGBoost and a Hybrid approach model. The Random Forest model demonstrates a training accuracy rate of 1.00 which decreases to 0.83 during test evaluations because of potential overfitting conditions. The XGBoost along with the Hybrid model demonstrate identical accuracy values of 0.85 in their training and test performances suggesting they possess stronger generalization capacities as well as model stability. Real-world deployment benefits extensively from XGBoost and Hybrid models since they demonstrate consistent performance for predicting patient data which has not been previously observed. The recommended drugs achieve superior results through better bias-variance control in the recommendations made by XGBoost and Hybrid when compared to Random Forest.

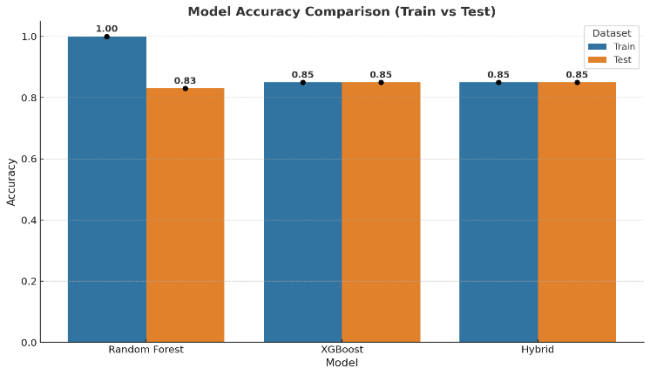

Fig. 5. Models' Accuracy analysis for both training and testing splits.

TABLE V. COMPARISON OF MODEL ACCURACY WITH DIFFERENT RESEARCH

Research	***Model used***	***Accuracy***
Mohapatra et al [8]	Logistic Regression	80.5%
Sae-Ang et. al [10]	Hybrid model using classifiers	74.6%
Reehan et. al. [11]	Naïve Bayes	84%
Our Approach	**Hybrid Ensemble Model**	**85%**

Table V shows that a machine learning model integrated into the personalized medicine recommendation system used patient data to forecast therapy with good results. The Random Forest model performed well during training but declined significantly in evaluation ratings, suggesting overfitting. XGBoost and Hybrid's 0.85 accuracy on both training and testing sets shows their generalizability and stability. Because medical personnel need accurate projections for unknown patient information, stability is crucial. Ensemble approaches through XGBoost processed complex non-linear healthcare data well. The research shows that the Hybrid and XGBoost models are suitable for personalized medicine recommendation systems.

V. Future Directions

The future scope of this study is to improve system prediction, this study will include more clinical, genetic, and behavioural data to increase individualization. Integration with real-time monitoring systems and EHRs allows adaptive drug recommendations. Deep learning and attention-based models should improve accuracy, especially for complex patient histories. Explainable AI solves class imbalance and interpretability, making the system more open and dependable for clinical use. Increasing the model's relevance across diseases and demographics could improve healthcare delivery.

VI. Conclusion

In conclusion, the application of machine learning for drug recommendation utilizes sophisticated algorithms and data-driven insights to enhance healthcare delivery. Utilizing predictive analytics and tailored interventions, machine learning-based systems can enhance patient care, elevate treatment outcomes, and transform healthcare methodologies. In terms of accuracy and generalization, ensemble techniques like XGBoost and Hybrid models surpass conventional approaches. Although the Random Forest model indicated indications of overfitting, the steady performance of XGBoost and the Hybrid model (both attaining 85% accuracy on training and test data) emphasizes their appropriateness for actual healthcare uses.

.

References

[1] "Disease Prediction and Drug Recommendation using Machine Learning", IJSDR - International Journal of Scientific Development and Research (www.IJSDR.org), ISSN:2455-2631, Vol.9, Issue 4, page no.1109 - 1115, April-2024, Available :https://ijsdr.org/papers/IJSDR2404160.pdf.

[2] Bhoomika Dubba, T. Sowmya Sri, M. Sowbhagya Pradhaini, DR. KC Sreedhar, “Drug Recommendation System Based on Sentiment Analysis of Drug Reviews Using Machine Learning”, IJFMR-International Journal for Multidisciplinary Research, E-ISSN: 2582-2160, Volume 6, Issue 3, May-June 2024 .

[3] Mariam Zomorodi, Ismail Ghodsollahee, Jennifer H Martin, Nicholas J Talley, Vahid Salari, Paweł Pławiak, Kazem Rahimi, U.R. Acharya, RECOMED: A comprehensive pharmaceutical recommendation system, Artificial Intelligence in Medicine, Volume 157, 2024, 102981, ISSN 0933-3657, https://doi.org/10.1016/j.artmed.2024.102981.

[4] Xiang Li, Shunpan Liang, Yu Lei, Chen Li, Yulei Hou, Dashun Zheng, and Tengfei Ma. 2024. CausalMed: Causality-Based Personalized Medication Recommendation Centered on Patient Health State. In Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (CIKM '24). Association for Computing Machinery, New York, NY, USA, 1276–1285. https://doi.org/10.1145/3627673.3679542

[5] Bak M, Madai VI, Fritzsche MC, Mayrhofer MT, McLennan S. You Can't Have AI Both Ways: Balancing Health Data Privacy and Access Fairly. Front Genet. 2022 Jun 13;13:929453. doi: 10.3389/fgene.2022.929453. PMID: 35769991; PMCID: PMC9234328.

[6] Komal Kumar, N., Vigneswari, D. (2021). A Drug Recommendation System for Multi-disease in Health Care Using Machine Learning. In: Hura, G.S., Singh, A.K., Siong Hoe, L. (eds) Advances in Communication and Computational Technology. ICACCT 2019. Lecture Notes in Electrical Engineering, vol 668. Springer, Singapore. https://doi.org/10.1007/978-981-15-5341-7_1.

[7] Mokhtari M, Khoshbakht S, Akbari ME, Moravveji SS. BMC3PM: bioinformatics multidrug combination protocol for personalized precision medicine and its application in cancer treatment. BMC Med Genomics. 2023 Dec 12;16(1):328. doi: 10.1186/s12920-023-01745-y. Erratum in: BMC Med Genomics. 2024 Jan 3;17(1):13. doi: 10.1186/s12920-023-01784-5. PMID: 38087279; PMCID: PMC10717810.

[8] Bhoi, S. (2020). PREMIER: Personalized REcommendation for Medical prescrIptions from Electronic Records. ArXiv.

[9] Mohapatra, M., Nayak, M., & Mahapatra, S. (2022). A machine learning based drug recommendation system for health care. *Graduate Research in Engineering and Technology (GRET)*, *1*(6).

[10] Al Mubasher, H., Doughan, Z., Sliman, L., Haidar, A. (2023). A Novel Neural Network-Based Recommender System for Drug Recommendation. In: Iliadis, L., Maglogiannis, I., Alonso, S., Jayne, C., Pimenidis, E. (eds) Engineering Applications of Neural Networks. EANN 2023. Communications in Computer and Information Science, vol 1826. Springer, Cham. https://doi.org/10.1007/978-3-031-34204-2_46.

[11] Sae-Ang, A., Chairat, S., Tansuebchueasai, N., Fumaneeshoat, O., Ingviya, T., & Chaichulee, S. (2023). Drug Recommendation from Diagnosis Codes: Classification vs. Collaborative Filtering Approaches. *International Journal of Environmental Research and Public Health*, *20*(1), 309. https://doi.org/10.3390/ijerph20010309.

[12] Reehan, P. M., Nikhilesh, N., Revanth, R., Chandra, V. V., & Suvanam, S. B. (2024, February). Disease prediction and medicine recommendation system. In AIP Conference Proceedings (Vol. 2742, No. 1). AIP Publishing, https://doi.org/10.1063/5.0184521.

[13] Ramachandran, R., Chinnammal, V., Malathy, K., Sailaja, K., & Shanmugam, S. (2024, May). An Improved Drug Recommendation System Using Artificial Intelligence Assisted Learning Methodology. In *2024 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)* (pp. 1-7). IEEE, 10.1109/ACCAI61061.2024.10602473.

[14] Thorat, S. A., Ashwini, G., & Seema, M. (2023, January). Survey on collaborative and content-based recommendation systems. In *2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT)* (pp. 1541-1548). IEEE, 10.1109/ICSSIT55814.2023.10061072.

[15] Machado, J., Rodrigues, C., Sousa, R., & Gomes, L. M. (2025). Drug–drug interaction extraction-based system: An natural language processing approach. *Expert Systems*, *42*(1), e13303, **https://doi.org/10.1111/exsy.13303**.

[16] KV, S. P. R., & Selvakumar, M. (2025). Personalized recommendation system to handle skin cancer at early stage based on hybrid model. *Network: Computation in Neural Systems*, 1-40, https://doi.org/10.1080/0954898X.2024.2449173.