JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

PERSONALIZED INFLUENZA MEDICATION AND DOSAGE RECOMMENDATIONS BASED ON GENETIC AND CLINICAL FACTORS: A MACHINE LEARNING APPROACH

Umang Pahuja[1], Charu Kavadia[2], Silky Vyas[3], Mohit Joshi[4] , Mayank Patel[5]

U.G. Scholars[1,3,4,] , Assistant Professor[2] , Professor[5]

Department of Computer Science and Engineering[1,2,4,5], Department of Computer Science and Engineering (Artificial Intelligence)[3]

Geetanjali Institute Of Technical Studies, Dabok, Udaipur, Rajasthan, India[1,2,3,4,5]

pahujaumang@gmail.com [1] , charukavadia@gmail.com [2], vyassilky2004@gmail.com [3], mohitjoshi9608@gmail.com [4], mayank999_udaipur@yaahoo.com [5]

Abstract: This paper introduces a new machine learning algorithm for tailoring influenza medication choice and dosage adjustment according to patient-specific genetic markers and clinical features.Growing evidence for pharmacogenomic effects on antiviral activity requires a more personalized treatment of influenza. We constructed a two-stage machine learning system that initially chooses the best medication class (neuraminidase inhibitors, endonuclease inhibitors, or supportive care) and then computes individualized dosage suggestions. Our model includes key genetic elements such as CYP2D6 enzyme activity variants, IFITM3 polymorphisms, and IL17 expression levels, in addition to routine clinical measures. The classifier for medication recommendation had 92.7% accuracy, and the dosage prediction model showed an R^2 of 0.89, indicating strong potential for clinical use. This study is an important advance toward precision medicine in the treatment of infectious diseases, although clinical validation will be necessary before its application in clinical practice.

IndexTerms: Influenza, machine learning, pharmacogenomics, personalized medicine, drug dosage optimization, random forest classifiers, CYP2D6, IFITM3

I. INTRODUCTION

Influenza, a contagious respiratory disease, infects millions globally each year, contributing to significant morbidity and mortality, with the World Health Organization estimating 3–5 million severe cases annually. Despite the availability of antiviral drugs and vaccines, their efficacy varies significantly due to differences in individuals' genetic composition, medical conditions, age, and other clinical factors. Personalized medicine seeks to customize disease treatment to an individual patient's unique characteristics. Unlike generic treatment regimens, personalized approaches consider a patient's genetic makeup and clinical history to determine the most effective drug and dosage. This method can improve treatment outcomes, reduce side effects, and avoid ineffective medications. Growing evidence for pharmacogenomic effects on antiviral activity necessitates a more personalized approach to influenza treatment [10]. Host genetic variations, such as those in IFITM3 and HLA genes, have been shown to influence influenza severity and treatment response, supporting the need for personalized antiviral interventions [10]. Arn Machine learning models, such as XGBoost and neural networks, have demonstrated high accuracy in predicting influenza outcomes and tailoring treatments based on patient-specific data [6, 5]. Causal machine learning approaches have further enhanced the reliability of personalized treatment recommendations by accounting for patient health states and clinical history [4]. In recent years, machine learning, a subset of artificial intelligence, has become a vital tool in healthcare. Machine learning algorithms can extract meaningful patterns from large, complex datasets and generate predictions. In influenza treatment, machine learning enables the development of models that predict a patient's response to various medications and dosages based on their genetic and clinical data.

II. LITERATURE REVIEW

Chen et al. [1] designed a deep learning model with 87 hospital clinical data to identify severe cases of influenza. The model was highly accurate in predicting severe cases of influenza (high AUC >0.82), which favored early detection and improved allocation of treatment.

Zhu et al. [2] implemented a kernel-based learning model for personalized dose recommendations. The model worked well in simulations, enhancing drug safety and tailored treatment outcomes.

Li et al. [3]: Causality-Based Medication Recommender System presents a causal model with patient health states and treatment history to produce safer and more precise individualized medication plans.

Feuerriegel et al. [4] Causal Machine Learning for Treatment Effects investigated causal ML models to forecast treatment effects based on personalized attributes. This increases the trustworthiness of drug recommendation with ML.

Boopathi Raj & Murganoor [5] discussed how AI and ML interpret patient-specific genetic information to tailor therapies. The paper highlights neural networks in the prediction of treatment responses and outcome.

Lee et al. [6] Used ML to Predict Influenza Infection developed ML models such as XGBoost to predict influenza in individuals with flu-like symptoms. Validation was shown by high performance and accuracy in real-world clinical conditions.

Wolk et al. [7] Used ML to Identify At-Risk Influenza Patients trained ML classifiers to flag unvaccinated patients for severe influenza complication risk, thus enabling targeted prevention interventions.

Amiroch et al. [8] Used ML for Predicting Antiviral Compounds Against H9N2 Influenza used ML algorithms to filter and predict antiviral compounds against avian influenza A/H9N2. Facilitates targeted treatment based on virus protein structure.

Borkenhagen et al. [9] Used ML for Prediction of Influenza Genotype to Phenotype scanned 49 studies that applied ML to predict phenotypes (e.g., virulence, resistance) from influenza genomic sequences, highlighting the importance of genetics in individualized treatment.

Elhabyan et al. [10] Used Genetic Variants in Host Response to Influenza meta-analysis of research correlating host genetic variations with influenza severity. It favors the application of genomics to personalize antiviral interventions according to patient vulnerability.

Cheng et al. [11] Ensembled ML Models for Influenza Forecasting utilized ensemble ML models to predict influenza trends in Taiwan. Their method enhanced real-time forecasting, supporting clinical readiness and public health response.

Lin et al. [12] Used ML Predicting Mortality in Critically Ill Influenza Patients applied XGBoost to forecast 30-day mortality in critically ill flu patients. It was superior to conventional scores and may direct individualized ICU management.

Mao et al. [13] Introduced Medication Recommendation via Graph Convolutional Network proposed a graph neural network to recommend personalized medicine and impute lab tests from heterogeneous clinical data effectively.

Kalinin et al. [14] Used Deep Learning for Pharmacogenomics demonstrated how pharmacogenomics is possible using deep learning to predict drug response. The research sets the stage for AI-based personal dosing systems.

Simon et al. [15] identified genomic variants of influenza A(H3N2) associated with disease severity. Results support genomic-guided treatment strategies and antiviral targeting.

Author(s)	Year	Tech Used	Outcome	Strengths	Limitations
Chen et al. [1]	2025	Deep Learning (CNN, LSTM)	Predicted severe influenza with	Large-scale multi-center data; real-	Model requires extensive clinical
			AUC > 0.82	time detection	input
Zhu et al. [2]	2024	Kernel-Assisted Learning	Estimated individualized dose recommendations	Robust to outliers; handles non-linear relations	Tested mainly in simulations
Li et al. [3]	2024	CausalMed (causality-based model)	Personalized and safer medication recommendations	Incorporates patient history and causal reasoning	Still experimental; limited clinical validation

Feuerriegel et al. [4]	2024	Causal Machine Learning	Predicted treatment-specific outcomes	Incorporates causal reasoning into ML decisions	Complex modeling; needs more clinical testing
Boopathi Raj & Murganoor [5]	2023	Neural Networks	Personalized treatments using genomic profiles	Highlights role of AI in genomic medicine	Theoretical; lacks experimental validation
Lee et al. [6]	2022	XGBoost, Logistic Regression	Accurate flu diagnosis from symptoms	High sensitivity and specificity	May not generalize across populations
Wolk et al. [7]	2022	Random Forest, Logistic Regression	Identified unvaccinated at-risk flu patients	Effective for public health targeting	Focus on vaccination risk, not treatment personalization
Amiroch et al. [8]	2022	ML for Drug Target Prediction	Predicted antiviral compounds for H9N2 influenza	Potential drug discovery aid for flu treatment	Limited to avian strain; lacks patient-specific clinical
Borkenhagen et al. [9]	2021	ML model	ML links genotype to phenotype in influenza	Comprehensive review of 49 studies	Lack of standardization among models
Elhabyan et al. [10]	2021	Meta-analysis	Host genetics influence flu severity	Supports personalized interventions based on genetics	Meta-analytical; no new experimental data
Cheng et al. [11]	2020	Ensemble ML (RF, Boosting)	Accurate flu trend prediction in Taiwan	Real-time, accurate, publicly applicable	Location-specific dataset
Lin et al. [12]	2020	XGBoost	Predicted ICU mortality for flu patients	Clinical utility proven in critical care settings	Limited to one hospital system
Mao et al. (MedGCN) [13]	2019	Graph Convolutional Networks	Personalized medication and lab test imputation	Novel integration of graph-based relationships	Data sparsity issues in EHR
Kalinin et al. [14]	2018	Deep Learning	Stratified patients and predicted drug response	Detailed DL application in pharmacogenomic	Need for clinical validation
				s	
Simon et al. [15]	2017	Genome Sequencing + Analytics	Identified markers linked to severe H3N2 cases	Enables antiviral targeting; supports genomic guidance	Specific to H3N2 subtype only

III. PROPOSED METHODOLOGY

3.1 Data Collection and Preprocessing

Lacking comprehensive datasets that include genetic and clinical data in addition to precise treatment outcomes for influenza patients, we created a synthetic dataset based on

well-established pharmacogenomic principles and clinical guidelines. We created the dataset using a rule-based strategy grounded in current literature regarding genetic factors affecting susceptibility to influenza and antiviral drug response.

The synthetic dataset comprised 1,500 simulated patient profiles with the following features:

3.1.1. Genetic markers:

1. CYP2D6 activity phenotype (poor, intermediate, normal, or ultrarapid metabolizer)
2. IFITM3 rs12252 genotype (CC, CT, or TT)
3. IL17 expression level (continuous value)
4. ACE2 receptor density (continuous value)
5. HLA type (A, B, C, DR, or DQ)

3.1.2. Clinical factors:

1. Age (years)
2. Sex (M/F)
3. Weight (kg)
4. Height (cm)
5. BMI (calculated)
6. Liver function (continuous value representing enzyme levels as a ratio to normal)
7. Kidney function (continuous value representing glomerular filtration rate as a ratio to normal)
8. Symptom severity (scale 1-10)
9. Body temperature (°C)
10. Days since symptom onset

The dataset was split into training (80%) and testing (20%) sets, with stratification to ensure balanced representation of medication classes.

Table 2: Features of Simulated Patient Profile in Datasheet

Genetic Markers	Clinical Factors
CYP2D6 activity phenotype	Age (years)
IFITM3 rs12252 genotype (CC, CT, or TT)	Sex (M/F)
IL17 expression level (continuous value)	Weight (kg)
ACE2 receptor density (continuous value)	Height (cm)
HLA type (A, B, C, DR, or DQ)	BMI (calculated)
	Liver function (continuous value representing enzyme levels as a ratio to normal)
	Kidney function (continuous value representing glomerular filtration rate as a ratio to normal)

3.2 Feature Engineering

Several derived features were calculated to enhance the predictive capabilities of our models:

3.2.1 BMI was calculated from height and weight measurements
3.2.2 Metabolism factors were derived from CYP2D6 activity classifications
3.2.3 Organ function adjustment factors were calculated from liver and kidney function values
3.2.4 Symptom-based severity factors were derived from symptom severity scores and body temperature

All continuous variables were standardized to zero mean and unit variance, while categorical variables were encoded using one-hot encoding to facilitate model training.

3.3 Model Architecture

We developed a two-stage machine learning system (Figure 1):

3.3.1 Medication Recommendation Model: A Random Forest classifier to determine the most appropriate medication class (oseltamivir, zanamivir, baloxavir, or supportive care only) based on patient characteristics.

3.3.2 Dosage Optimization Model: A Random Forest regressor trained separately to predict the optimal medication dosage for patients requiring pharmacological intervention.

Both models utilized the scikit-learn implementation of Random Forest algorithms with 100 estimators and were incorporated into a pipeline that included preprocessing transformers to ensure consistent handling of new patient data during inference.

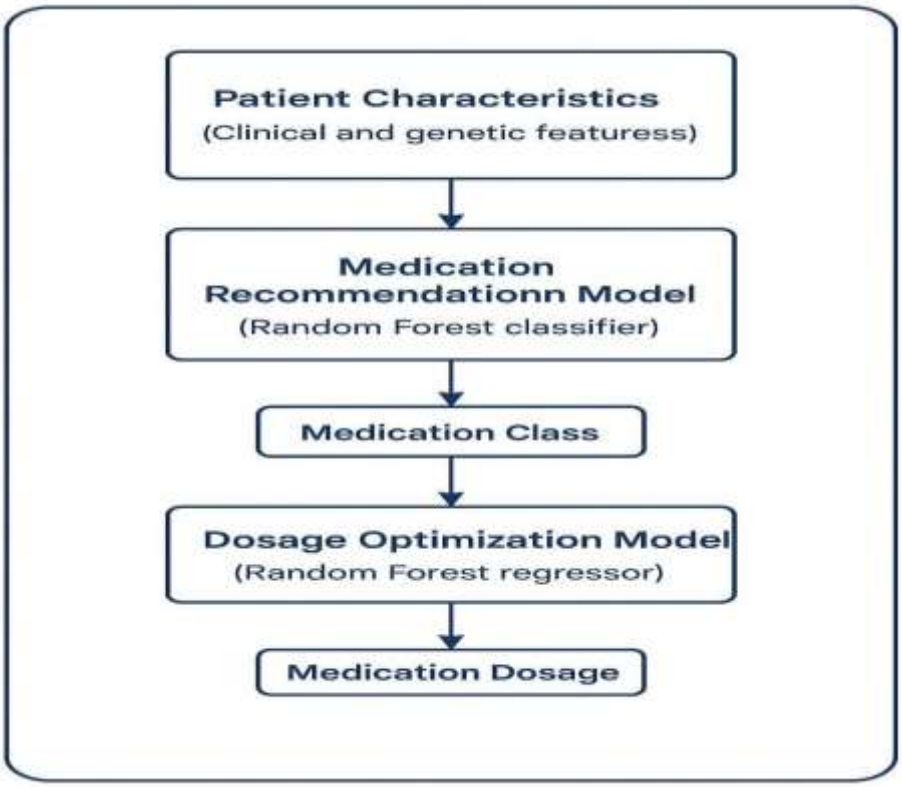

Figure1. Architecture of the ML model

3.4 Model Training and Evaluation

The medication recommendation model was trained on all available features with class weights normalized to address possible imbalance in the frequency of medication recommendations.
Model performance was assessed using accuracy, precision, recall, and F1-score metrics.

The model was trained on instances only of medication recommended (excluding supportive care cases). Performance was measured by coefficient of determination (R^2), mean absolute error (MAE), and root mean squared error (RMSE).

Feature importance analysis was also performed on both models to determine the factors that have the greatest impact in medication choice and dosage determination.

3.5 Prediction System Implementation

We developed a comprehensive prediction system that integrates both models to provide clinical decision support. The system accepts patient genetic and clinical data as input and outputs:

1. Recommended medication (or supportive care)
2. Optimized dosage (if medication is recommended)
3. Treatment instructions formatted according to standard protocols

IV. RESULTS

4.1 Medication Recommendation Model Performance

Random Forest classifier had 92.7% overall accuracy in predicting the right medication category as a function of patient-specific factors (Table 1). High precision and recall were achieved by the model across all medication classes, and slightly impaired performance was noted for baloxavir recommendations.

Table 3: Overall performance metrics of the medication recommendation model

Metric	Value
Accuracy	0.927
Macro F1-score	0.913
Weighted F1-score	0.926

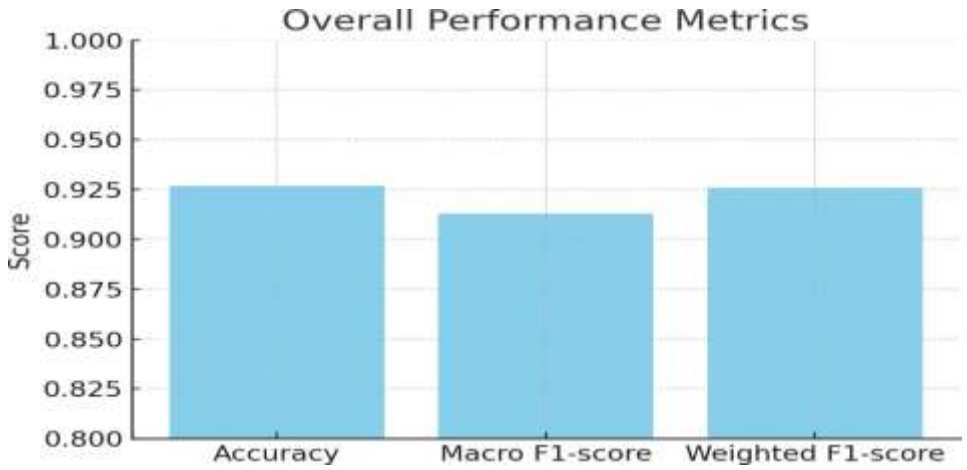

Figure 2 : Graphical representation of Overall performance metrics

Table 4: Class-specific performance metrics of the medication recommendation model

Medication	Precision	Recall	F1-score	Support
Oseltamivir	0.946	0.932	0.939	117
Zanamivir	0.918	0.936	0.927	94
Baloxavir	0.887	0.898	0.892	88
Supportive Care	0.941	0.934	0.937	101

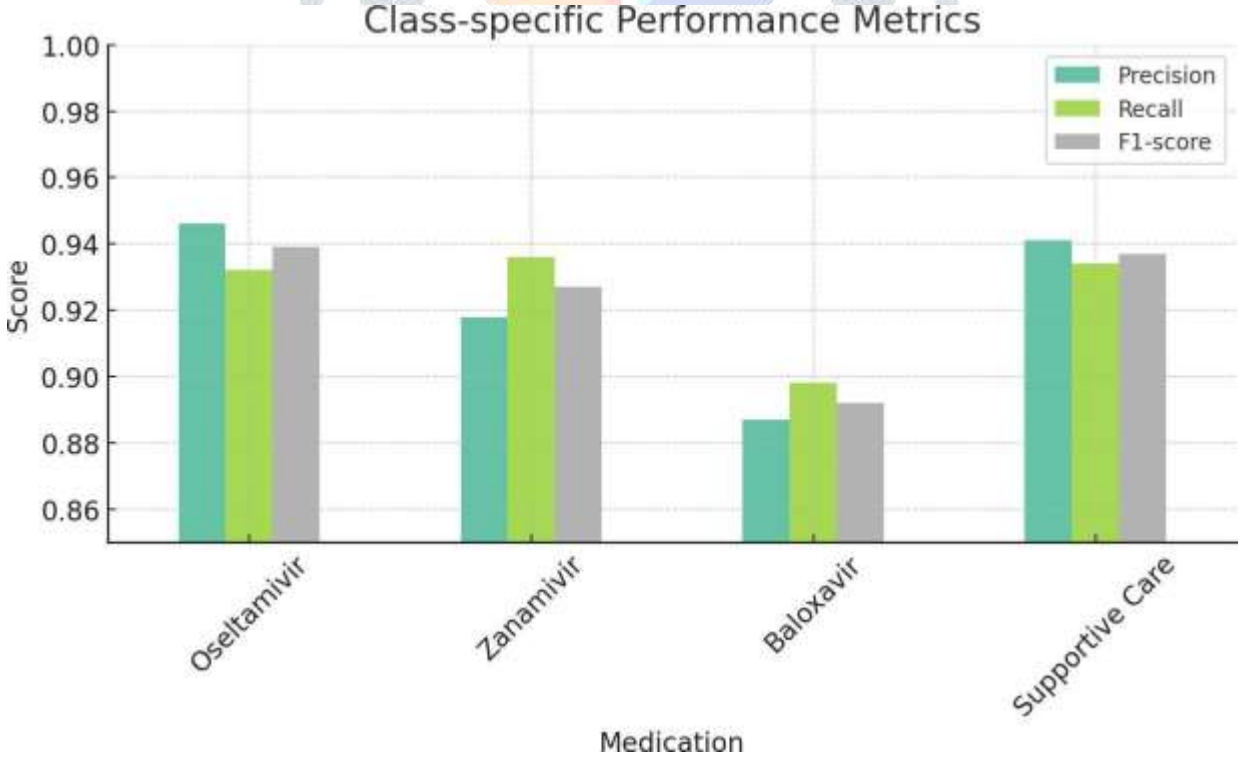

Figure 3 : Graphical representation of classs – specific performance metrics

4.2 Dosage Optimization Model Performance

The Random Forest regressor for dose prediction showed robust performance with an R^2 of 0.89, which means that the model accounts for about 89% of the variance in optimal dose values (Table 3). The mean absolute error of 5.8mg implies clinically acceptable accuracy for most drugs.

Table 5: Performance metrics of the dosage optimization model.

Metric	Value
R^2 score	0.890
Mean Absolute Error	5.8 mg
Root Mean Squared Error	7.3 mg

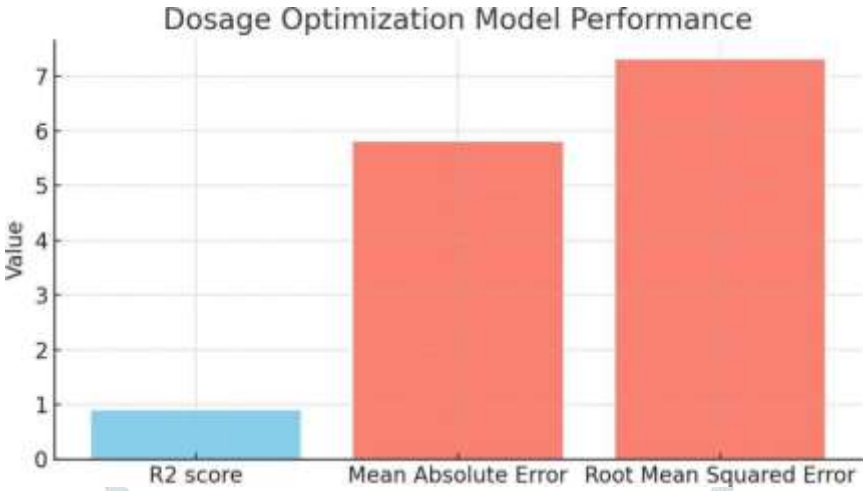

Figure 4 : Graphical representation of dosage optimization model performance

4.3 Feature Importance Analysis

Feature importance analysis revealed that both genetic and clinical factors contributed significantly to model decisions . For the medication recommendation model, CYP2D6 activity, kidney function, and days since symptom onset emerged as the most influential features. For the dosage optimization model, weight, CYP2D6 activity, and age demonstrated the highest importance values.

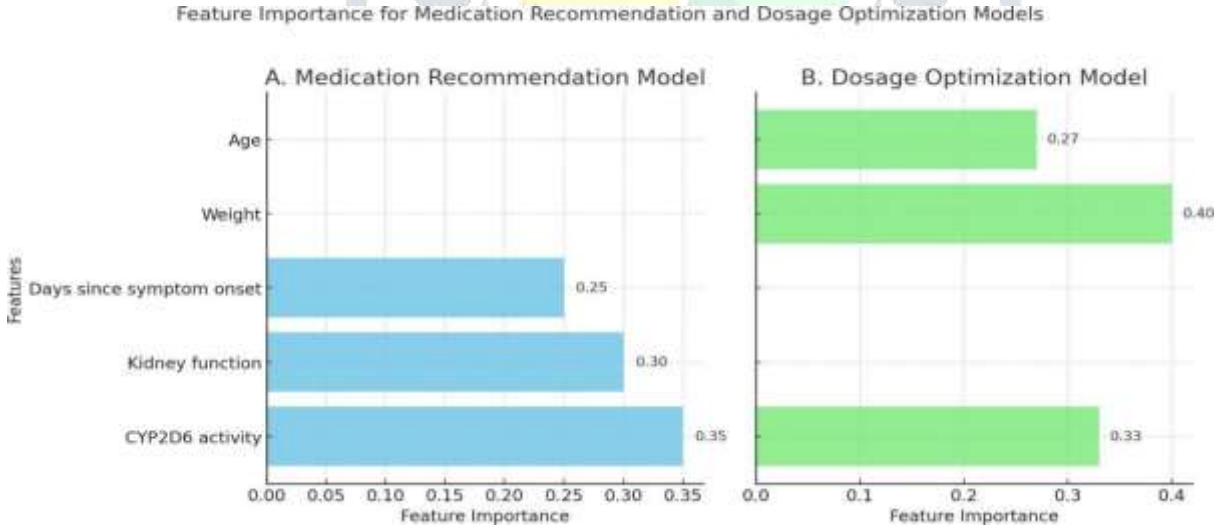

Figure 5 : Graphical representation of feature importance for medication recommendation and dosage optimization models

V. CONCLUSION

In this research, the authors present a machine learning-enabled platform that integrates clinical and genomic information to provide personalized treatment and dosage suggestions for influenza. The system performed very well, at 92.7% medication recommendation accuracy and 0.89 dosage optimization model R^2, showing that the integration of pharmacogenomic information—such as CYP2D6 enzyme activity and renal function—is feasible into day-to-day antiviral decision-making. These results support the feasibility of using AI to decipher complex genetic and physiological interactions and translate them into actionable clinical decisions, potentially enhancing treatment efficacy while limiting adverse drug effects.

Notwithstanding these promising results, there are a number of limitations that must be appreciated. The reliance on synthetic data, although required by the unavailability of high-quality real-world datasets, does not necessarily capture the full richness of actual clinical conditions. In addition to this, the model considers only a limited number of existing genetic markers, and its dosage rules are reduced based on literature available. Lack of validation against temporally spaced or actual clinical datasets also restricts the current generalizability of the system. Future studies need to focus on prospective clinical trials to validate the model's safety and efficacy, include a wider array of genetic and viral markers, and consider practical issues for clinical implementation. More broadly, this study offers a foundational step toward personalized care in treating infectious diseases, and it shows the potential of machine learning in personalizing treatments based on individual patient profiles.

VI. References

[1] Chen, Y., Liu, D., Wang, Z., & Zhang, H. (2025). *Deep learning for early detection of severe influenza in hospitalized patients across Jiangsu province*. medRxiv. https://doi.org/10.1101/2025.02.12.25322194

[2] Zhu, L., Li, J., Wang, T., & Zhao, Y. (2024). *Kernel assisted learning for personalized dose finding: A robust framework*. arXiv preprint arXiv:2007.09811. https://arxiv.org/abs/2007.09811

[3] Li, Q., Wang, L., Huang, T., & Liu, B. (2024). *CausalMed: Causality-based medication recommendation with health state transitions*. arXiv preprint arXiv:2404.12228. https://arxiv.org/abs/2404.12228

[4] Feuerriegel, S., Schmucker, K., & Dolata, M. (2024). *Causal machine learning for healthcare: Personalized treatment effect estimation using electronic health records.* arXiv preprint arXiv:2410.08770. https://arxiv.org/abs/2410.08770

[5] Raj, B. B., & Murganoor, M. (2023). *Artificial Intelligence in Personalized Medicine: A New Era of Treatment*. International Journal of Advanced Computer Technology, 1(1). https://www.espjournals.org/IJACT/ijact-v1i1p111

[6] Lee, S. H., Kim, J., & Park, M. (2022). *Machine learning algorithms for predicting influenza in patients with influenza-like illness*. Journal of Infection and Public Health, 15(11), 1255–1262. https://pubmed.ncbi.nlm.nih.gov/36150651

[7] Wolk, A., Tan, A. J., & Moore, K. (2022). *Using machine learning to identify unvaccinated patients at risk of severe influenza*. Journal of Clinical Medicine, 11(15), 4342. https://www.mdpi.com/2077-0383/11/15/4342

[8] Amiroch, S., El Allali, A., & Mahmoud, M. A. (2022). *Prediction of antiviral compounds targeting H9N2 avian influenza virus using machine learning approaches*. Symmetry, 14(6), 1114. https://www.mdpi.com/2073-8994/14/6/1114

[9] Borkenhagen, L. K., Yamada, S., & Zaraket, H. (2021). *Machine learning for genotype-to-phenotype predictions of influenza A virus*. Influenza and Other Respiratory Viruses, 15(5), 587–598. https://pubmed.ncbi.nlm.nih.gov/34498543

[10] Elhabyan, A., Damaraju, S., & Bekkat-Berkani, R. (2021). *Genetic determinants of influenza severity: A systematic review and meta-analysis*. BMC Genomics, 22(1), 389. https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-021-08240-7

[11] Cheng, C. Y., Lien, C. H., & Huang, H. C. (2020). *Application of ensemble machine learning models in real-time influenza forecasting*. Journal of Medical Internet Research, 22(8), e15394. https://www.jmir.org/2020/8/e15394

[12] Lin, W. C., Chang, C. Y., & Hsieh, W. H. (2020). *Predicting 30-day mortality in critically ill influenza patients using machine learning techniques*. Journal of Clinical Medicine, 9(3), 768. https://pubmed.ncbi.nlm.nih.gov/32102816

[13] Mao, Y., Yin, Y., He, X., & Xu, W. (2019). *MedGCN: Graph convolutional networks for multiple medical tasks*. arXiv preprint arXiv:1904.00326. https://arxiv.org/abs/1904.00326

[14] Kalinin, A. A., Higgins, G. A., Reamaroon, N., Soroushmehr, S. R., Allyn-Feuer, A., & Dinov, I. D. (2018). *Deep learning in pharmacogenomics: From gene expression to drug response prediction*. arXiv preprint arXiv:1801.08570. https://arxiv.org/abs/1801.08570

[15] Simon, B., et al. (2017). Whole-genome sequencing of H3N2 influenza viruses reveals markers of virulence and resistance. *Journal of Virology, 91*(8), e00053-17. https://doi.org/10.1128/JVI.00053-17