Journal of Artificial Intelligence and Metaheuristic

♦ ISSN: 2833-5597 Vol: 12, No. 01, 2027 ♦ PP. 11 – 17

♦ Article DOI: https://doi.org/10.54216/JAIM.120102

REVIEW ARTICLE

Comparative Explainable Machine Learning Framework for Multiclass Drug Classification

Ahmed Ashraf Abdelfatah[1,*]

[1] Department of Clinical Pharmacy and Pharmacy Practice, Faculty of Pharmacy, Mansoura University, Mansoura 35516, Egypt

Email: ahmedmohamedashraf@std.mans.edu.eg

Received: May 31, 2026 Revised: July 14, 2026 Accepted: September 06, 2026 ⋆ Corresponding author

ABSTRACT

Accurate drug classification is important for developing intelligent healthcare systems and supporting reliable medication-related decision-making. The relationships between patient characteristics and prescribed drug categories may be nonlinear and difficult to identify using conventional analytical procedures. This study presents a machine learning framework for multiclass drug classification using demographic and clinical patient attributes. The framework includes data cleaning, exploratory data analysis, numerical and categorical feature transformation, model training, comparative evaluation, and explainable artificial intelligence analysis. Nine classification models were investigated, including Random Forest, support vector machine, LightGBM, CatBoost, multilayer perceptron, one-dimensional convolutional neural network, long short-term memory network, gated recurrent unit network, and TabNet. The models were evaluated using accuracy, weighted precision, weighted recall, weighted F1-score, macro F1-score, and balanced accuracy. Random Forest achieved the best overall balance between predictive performance and interpretability, obtaining an accuracy of 97.50%, a macro F1-score of 98.51%, and a balanced accuracy of 98.18%. SHAP analysis identified the sodium-to-potassium ratio, blood-pressure categories, and age as the most influential variables affecting the classification decisions. The obtained results demonstrate that an explainable machine learning framework can provide accurate and transparent drug classification using a limited number of patient-related variables.

Keywords: Drug classification ▪ Machine learning ▪ Random Forest ▪ Deep learning ▪ Explainable artificial intelligence ▪ SHAP ▪ Healthcare decision support

1. INTRODUCTION

Drug classification is an important application of artificial intelligence in pharmaceutical and healthcare systems because it supports the organization of drug-related information and assists in identifying suitable medication categories according to patient characteristics. The increasing availability of clinical and pharmaceutical data has created a need for computational methods capable of processing heterogeneous variables and detecting relationships that may be difficult to identify through conventional analysis. Artificial intelligence has consequently become an influential component of drug development, drug repositioning, and clinical decision-support systems [1, 2, 3].

Machine learning provides an effective framework for drug classification by learning relationships between patient characteristics and predefined drug classes. Patient records may contain numerical attributes, such as age and sodium-to-potassium ratio, together with categorical attributes, such as sex, blood-pressure level, and cholesterol category. These

variables may interact in complex and nonlinear ways. Machine learning classifiers can model such interactions and generate automated multiclass predictions without requiring manually defined decision rules. Previous studies have demonstrated the value of machine learning in drug–target interaction analysis and chemoinformatics applications [4, 5].

Despite its potential, patient-based drug classification remains challenging because predictive reliability may be affected by limited sample size, unequal class frequencies, heterogeneous feature types, nonlinear interactions, and inappropriate model selection. Models that achieve high overall accuracy may still perform inconsistently across individual classes when the dataset is imbalanced. Pharmacovigilance studies also show that pharmaceutical datasets may contain complex associations and heterogeneous observations that influence the reliability of computational predictions [6, 7].

Interpretability is another important requirement in healthcare-related applications. Predictive performance alone is insufficient because users may need to understand which patient attributes influence the model output. Explainable artificial intelligence can address this limitation by quantifying feature contributions and identifying the variables that most strongly affect classification decisions.

This study develops an explainable machine learning framework for multiclass drug classification using demographic and clinical patient attributes. The methodology includes data cleaning, stratified partitioning, feature transformation, class-weight calculation, training of conventional machine learning and neural models, comparative evaluation, and SHAP-based interpretation. The main contributions are the evaluation of nine classification models under a consistent procedure, the use of class-sensitive performance metrics, and the interpretation of the selected model using global SHAP feature importance.

2. LITERATURE REVIEW

Artificial intelligence and machine learning have become increasingly important in pharmaceutical research because they provide computational mechanisms for examining complex drug-related information and supporting evidence-based decisions. Existing research covers pharmaceutical formulation, drug-delivery systems, pharmacokinetic classification, therapeutic applications, and medication response. However, these applications differ from patient-oriented drug classification, in which a predictive model determines an appropriate drug category using demographic and clinical variables.

Research on pharmaceutical formulation has emphasized the importance of categorizing drugs and delivery systems according to their physical, chemical, and therapeutic properties. Chelliah *et al.* reviewed liposomal drug-delivery systems and classified them according to composition, therapeutic applications, and practical limitations [8]. Their work demonstrates the value of structured pharmaceutical categorization, although the classification is primarily based on formulation characteristics rather than patient-level clinical observations.

Uttreja *et al.* examined self-emulsifying drug-delivery systems and discussed their transition from liquid formulations to solid dosage forms [9]. Their review considered formulation components, characterization methods, pharmaceutical applications, and future development directions. This work shows how systematic classification can organize complex formulation information, but it does not directly address automated prediction of drug classes from patient-specific attributes.

Drug classification is also associated with pharmacokinetic behavior and therapeutic decision-making. Simitopoulos *et al.* investigated the relationship between drug dissolution and the Biopharmaceutics Classification System and explained how solubility, permeability, and dissolution characteristics contribute to pharmaceutical categorization [10]. In contrast to rule-based pharmaceutical systems, machine learning classification learns decision boundaries directly from observed patient records.

Therapeutic studies further indicate that medication selection can be informed by patient-specific clinical and demographic data. Benrimoh *et al.* developed a deep learning model to estimate remission probabilities across ten pharmacological treatments for major depression using clinical and demographic variables [11]. Their study illustrates how predictive models can support personalized treatment selection by comparing expected outcomes across medication options. Evaluating interacting patient characteristics manually can become difficult when several numerical and categorical variables are considered simultaneously.

Traditional machine learning algorithms are suitable for structured medical datasets because they can learn nonlinear boundaries from tabular variables. Tree-based ensemble methods can capture complex feature interactions and are generally effective when datasets contain numerical and encoded categorical features. Kernel-based classifiers can generate flexible decision boundaries, while neural networks can learn complex representations from transformed clinical inputs.

Deep learning methods, including convolutional and recurrent architectures, have been widely applied to sequential, spatial, and high-dimensional medical data. Their use with small tabular datasets requires careful evaluation because the number of available samples may be insufficient for estimating a large number of parameters. Therefore, direct comparison between conventional machine learning, ensemble learning, and neural models is important for determining which architectural assumptions are most suitable for patient-oriented drug classification.

Model interpretability is particularly important in medical and pharmaceutical applications. Feature-importance analysis can identify the variables that contribute most strongly to predictive decisions, while local and global explanation methods can determine whether a model relies on clinically meaningful information. SHAP provides a unified approach for estimating feature contributions and ranking numerical and encoded categorical variables according to their average influence on model predictions.

Overall, previous research confirms the importance of structured pharmaceutical categorization and computational analysis. Nevertheless, patient-based multiclass drug classification requires a framework that combines consistent preprocessing, multiple classifier families, class-sensitive evaluation, and model interpretation. The present study addresses this requirement by comparing conventional machine learning, ensemble, and neural models and interpreting the selected

classifier using SHAP analysis.

3. MATERIALS AND METHODS

The proposed framework consists of dataset inspection, preprocessing, exploratory data analysis, multiclass model development, performance evaluation, and model interpretation. All classifiers were trained and evaluated using the same cleaned dataset and the same stratified training and testing subsets to maintain a consistent comparison.

3.1 Dataset and Prediction Task

The study used a drug classification dataset containing 200 patient records. Each record includes five predictor variables: age, sex, blood-pressure category, cholesterol category, and sodium-to-potassium ratio. The target variable contains five prescription classes: DrugY, drugX, drugA, drugB, and drugC. The prediction problem was formulated as a supervised multiclass classification task in which one drug label was assigned to each patient record. The predictor variables represent heterogeneous data types. Age and sodium-to-potassium ratio are numerical variables, whereas sex, blood pressure, and cholesterol are categorical variables. Consequently, separate transformation procedures were required before model training. The use of a unified preprocessing pipeline ensured that transformations were estimated using the training subset and then applied consistently to the testing subset.

3.2 Data Cleaning and Preprocessing

The dataset was inspected before model development. Column names were standardized, and unnecessary spaces were removed from categorical values. Numerical columns were converted into numeric format so that invalid strings could be identified. Missing observations, invalid categories, and duplicate records were removed. Validity conditions were applied to the numerical features. Age values were retained when they were greater than zero and less than or equal to 120, while the sodium-to-potassium ratio was required to be greater than zero. Categorical observations were retained only when their values belonged to the defined categories of the corresponding variables. These procedures prevented malformed records from influencing model training. The cleaned dataset was divided into training and testing subsets using an 80:20 stratified split with a random seed of 42. Stratification preserved the approximate class proportions in both subsets. The target drug classes were transformed into numerical labels using label encoding. Balanced class weights were calculated from the training labels to reduce the influence of unequal class frequencies. Numerical features were imputed using the median of the corresponding training variable and standardized using `StandardScaler`. For a numerical feature x, standardization was performed as

$$z = \frac{x - \mu}{\sigma}, \tag{1}$$

where μ and σ denote the mean and standard deviation estimated from the training subset. Standardization placed numerical variables on comparable scales and reduced the effect of differences in their original ranges. Categorical variables were imputed using the most frequent training category and transformed using one-hot encoding. One-hot encoding converted each categorical state into a separate binary feature. The preprocessing pipeline was fitted exclusively using the training subset to prevent information leakage and was subsequently used to transform the testing subset.

3.3 Exploratory Data Analysis

Exploratory data analysis was conducted to examine the target-class distribution and identify important relationships between the patient characteristics and the prescribed drug categories. The analysis focused on class frequency, the distribution of the sodium-to-potassium ratio, and the composition of drug prescriptions within the blood-pressure categories. Figure 1 presents the number of patients associated with each drug class. DrugY is the largest class with 91 patients, followed by drugX with 54 patients and drugA with 23 patients. DrugB and drugC each contain 16 patients. The observed differences confirm that the target distribution is imbalanced and justify the calculation of balanced class weights during model development. Fig. 1 presents the number of patients associated with each drug class. DrugY is the largest class with 91 patients, followed by drugX with 54 patients and drugA with 23 patients. DrugB and drugC each contain 16 patients. The observed differences confirm that the target distribution is imbalanced and justify the calculation of balanced class weights during model development.

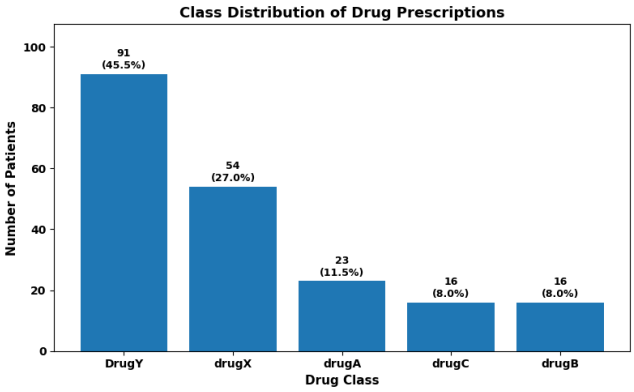

Figure 1. Distribution of patients across the five drug prescription classes.

Fig. 2 illustrates the sodium-to-potassium ratio across the five prescription classes. DrugY exhibits substantially higher values and greater variability than the remaining categories. In comparison, drugX, drugA, drugB, and drugC are concentrated within lower and narrower ranges. This separation suggests that the sodium-to-potassium ratio provides strong discriminatory information for distinguishing DrugY from the remaining classes.

Fig. 3 shows the proportional composition of drug classes within the LOW, NORMAL, and HIGH blood-pressure categories. DrugY represents a substantial proportion of prescriptions in the LOW and HIGH groups, whereas drugX dominates the NORMAL group. DrugA and drugB occur within the HIGH category, while drugC occurs within the LOW category. These patterns indicate that blood-pressure status provides useful information for separating the prescribed drug categories.

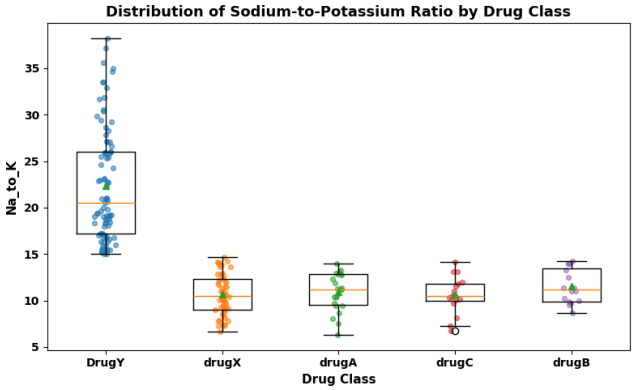

Figure 2. Distribution of the sodium-to-potassium ratio across the five drug prescription classes.

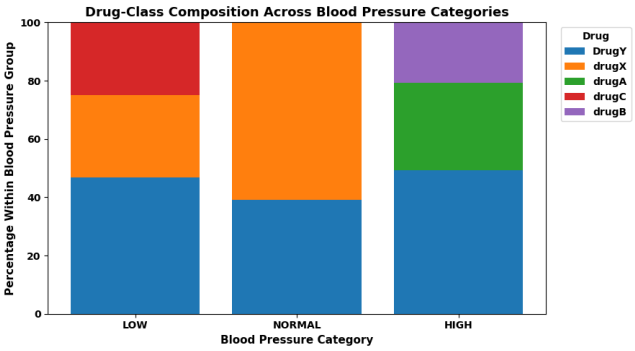

Figure 3. Percentage composition of drug prescription classes within each blood-pressure category.

3.4 Machine Learning Models

Nine classification models were evaluated: Random Forest, support vector machine with a radial basis function kernel, LightGBM, CatBoost, multilayer perceptron, one-dimensional convolutional neural network, long short-term memory network, gated recurrent unit network, and TabNet. The selected models represent ensemble learning, kernel-based learning, boosting, fully connected neural learning, convolutional learning, recurrent learning, and attention-based tabular learning. Random Forest was constructed using 300 decision trees. Each tree was trained using a sampled subset of the training data, and the final prediction was determined by aggregating the individual tree outputs. The ensemble structure reduces dependence on a single decision tree and allows nonlinear relationships and feature interactions to be modeled. The support vector machine used a radial basis function kernel, a regularization parameter of 1.0, automatic gamma scaling, probability estimation, and balanced class weights. The radial basis function kernel enabled nonlinear separation between the drug categories in the transformed feature space. LightGBM and CatBoost were trained using 300 estimators or iterations and a learning rate of 0.03. Both models use sequential boosting, in which new learners are added to correct errors produced by earlier learners. CatBoost also provides specialized treatment of categorical information, whereas LightGBM is designed for computationally efficient tree construction. The multilayer perceptron contained two hidden layers with 64 and 32 neurons. The one-dimensional convolutional neural network used convolutional layers with 32 and 64 filters to learn local relationships among the transformed feature values. The long short-term memory and gated recurrent unit models each contained 64 recurrent units. Although recurrent networks are commonly designed for sequential data, they were included to examine their performance on the transformed patient feature representation. TabNet used decision and attention dimensions of 16, three decision steps, a gamma value of 1.3, and a sparsity coefficient of 10^{-4}. Its attention mechanism performs sequential feature selection during the decision process and is specifically designed for tabular data.

3.5 Training Procedure

The neural models were trained using the Adam optimizer with a learning rate of 0.001 and sparse categorical cross-entropy as the loss function. Training was conducted for a maximum of 150 epochs. Early stopping was applied to terminate training when validation performance no longer improved, thereby reducing unnecessary iterations and limiting overfitting. Learning-rate reduction was also used when the validation performance reached a plateau. The same testing subset was used to evaluate all trained classifiers. Model comparison was based only on predictions generated for records that were not used during model fitting. This procedure ensured that the reported results represented predictive behavior on unseen samples.

3.6 Evaluation Metrics

The models were evaluated using accuracy, weighted precision, weighted recall, weighted F1-score, macro F1-score, and balanced accuracy. Accuracy measures the proportion of correctly classified testing records and is defined as

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}}. \tag{2}$$

For each class, precision measures the proportion of samples predicted as that class that were correctly classified, while recall measures the proportion of actual class samples that were correctly identified. The F1-score is the harmonic mean of precision and recall:

$$F1 = 2\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{3}$$

Weighted metrics account for class frequency by weighting the contribution of each drug class according to its number of samples. Macro F1 assigns equal importance to all classes by averaging their class-specific F1-scores without considering class size. Balanced accuracy similarly averages the recall values across the five classes. These class-sensitive measures are important because the dataset contains unequal target frequencies. Confusion matrices and classification reports were generated to inspect class-specific predictions. SHAP analysis was applied to Random Forest to determine the global contribution of each numerical and one-hot encoded feature. Global feature importance was calculated using the mean absolute SHAP value across the analyzed samples.

4. RESULTS AND DISCUSSION

This section presents and analyzes the experimental results obtained from the evaluated classification models. The discussion focuses on predictive performance, comparative model

behavior, and the interpretation of the selected classifier using SHAP analysis.

4.1 Classification Performance

The evaluated classifiers were compared using the six performance measures described in the methodology. Table 1 presents the complete results for all nine models. Random Forest achieved the best overall balance between predictive effectiveness and interpretability, obtaining an accuracy of 97.50%, weighted precision of 97.63%, weighted recall of 97.50%, weighted F1-score of 97.47%, macro F1-score of 98.51%, and balanced accuracy of 98.18%. CatBoost and 1D-CNN reached the same reported metric values as Random Forest. However, Random Forest was selected for the final interpretation because it provided strong predictive performance through a comparatively direct ensemble structure and was compatible with the implemented SHAP analysis. LightGBM achieved 95.00% accuracy, while MLP achieved 92.50%. SVM-RBF and TabNet each obtained 87.50% accuracy, although their class-sensitive measures differed. LSTM and GRU produced the lowest overall results. The difference between accuracy and macro F1 is especially important for the lower-performing recurrent models. LSTM achieved 52.50% accuracy but only 25.91% macro F1, while GRU achieved 50.00% accuracy and 24.42% macro F1. These results indicate that the recurrent models did not classify all drug categories consistently and that their overall accuracy was influenced by the unequal class distribution. The stronger performance of the tree-based models can be associated with the structure of the dataset. The input consists of a small number of tabular numerical and categorical attributes, and the total number of records is limited. Tree ensembles can model threshold-based decisions and nonlinear feature interactions without estimating the large number of parameters required by recurrent neural networks.

4.2 Comparative Analysis of the Models

Figure 4 compares the five strongest classifiers using accuracy, macro F1-score, and balanced accuracy. Random Forest, CatBoost, and 1D-CNN achieved identical reported values across the three displayed metrics. Random Forest obtained 97.50% accuracy, 98.51% macro F1-score, and 98.18% balanced accuracy. The high macro F1-score indicates reliable behavior across the five classes rather than performance dominated only by the largest category. LightGBM achieved 95.00% accuracy, 93.43% macro F1-score, and 94.18% balanced accuracy. MLP achieved lower accuracy than LightGBM but obtained a balanced accuracy of 96.67%, indicating relatively strong average recall across the drug classes. Considering predictive performance, model structure, and explainability, Random Forest was retained as the selected classifier. Fig. 4 compares the five best-performing classification models using accuracy, macro F1-score, and balanced accuracy.

4.3 Model Interpretability Using SHAP

SHAP analysis was used to determine which transformed input variables had the greatest global influence on the Random Forest predictions. Figure 5 ranks the features according to their mean absolute SHAP values. A larger mean absolute value indicates that a feature produced a greater average change in the model output across the analyzed patient

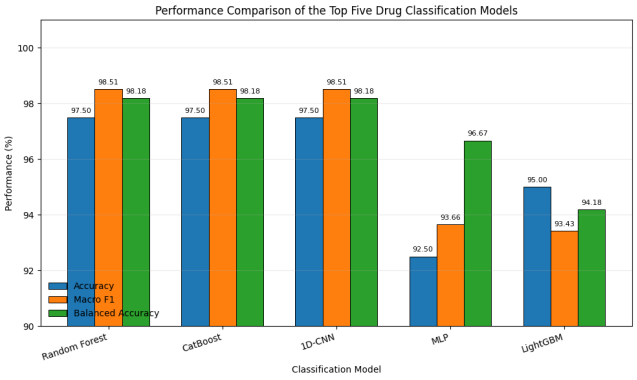

Figure 4. Comparison of the five best-performing classification models using accuracy, macro F1-score, and balanced accuracy.

records. The sodium-to-potassium ratio was the most influential variable, with a mean absolute SHAP value of 0.168094. The importance of this variable agrees with the exploratory boxplot, which showed a clear separation between DrugY and the remaining prescription categories. The `BP_HIGH` feature ranked second with a value of 0.095603, demonstrating the importance of blood-pressure status in differentiating the drug classes. Age produced a mean absolute SHAP value of 0.056074. The remaining blood-pressure features, `BP_LOW` and `BP_NORMAL`, achieved values of 0.044598 and 0.039046, respectively. The cholesterol features had smaller contributions, with values of 0.023240 for `Cholesterol_NORMAL` and 0.022412 for `Cholesterol_HIGH`. The encoded sex features produced the lowest global importance values.

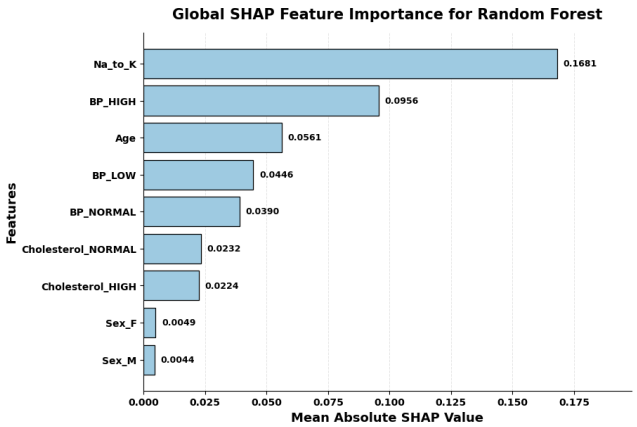

Figure 5. Global SHAP feature importance for the Random Forest classifier based on mean absolute SHAP values.

4.4 Discussion

The obtained results demonstrate that strong drug classification performance can be achieved using a limited number of demographic and clinical attributes. Random Forest, CatBoost, and 1D-CNN achieved the highest reported testing metrics. Nevertheless, Random Forest offers an advantageous balance because its ensemble decisions can be interpreted using feature-attribution methods while retaining high class-sensitive performance. The results also demonstrate that architectural complexity does not necessarily improve classification performance when the dataset is small and tabular. LSTM and GRU were originally designed to model temporal dependencies, whereas the patient records used in

Table 1. Performance comparison of all evaluated drug classification models.

Model	Accuracy (%)	Precision (%)	Recall (%)	Weighted F1 (%)	Macro F1 (%)	Balanced Accuracy (%)
Random Forest	97.50	97.63	97.50	97.47	98.51	98.18
CatBoost	97.50	97.63	97.50	97.47	98.51	98.18
1D-CNN	97.50	97.63	97.50	97.47	98.51	98.18
LightGBM	95.00	95.76	95.00	95.01	93.43	94.18
MLP	92.50	93.89	92.50	92.55	93.66	96.67
SVM-RBF	87.50	90.36	87.50	87.30	88.66	94.44
TabNet	87.50	88.73	87.50	86.29	85.00	85.78
LSTM	52.50	61.89	52.50	42.08	25.91	27.64
GRU	50.00	53.14	50.00	37.04	24.42	28.48

this study do not represent natural time sequences. Their lower macro F1 and balanced accuracy values indicate that their recurrent structures were not well matched to the characteristics of the prediction problem. The performance of SVM-RBF and TabNet was lower than that of the leading ensemble models. SVM-RBF achieved a balanced accuracy of 94.44%, suggesting that it retained useful class-level discrimination despite its lower overall accuracy. TabNet obtained lower macro F1 and balanced accuracy values, which may reflect the difficulty of training an attention-based tabular architecture using only a limited number of records. The SHAP findings strengthen the interpretation of the selected classifier. The sodium-to-potassium ratio, blood-pressure categories, and age were the dominant variables affecting the model decisions. These variables also exhibited visible relationships with the drug classes during exploratory analysis. The agreement between the exploratory patterns and the SHAP ranking indicates that Random Forest relied primarily on informative patient attributes rather than on weakly encoded variables. Although the results are promising, they should be interpreted in relation to the dataset size. The testing subset contains a limited number of records, and a small number of changed predictions can affect the reported percentages. Therefore, the present findings demonstrate the effectiveness of the implemented framework on the considered dataset but do not replace external clinical validation.

5. CONCLUSION

This paper presented an explainable machine learning framework for multiclass drug classification using demographic and clinical patient attributes. Nine conventional machine learning, ensemble, and neural classifiers were trained and evaluated under a consistent preprocessing and testing procedure. Random Forest provided the most suitable balance between predictive performance and interpretability, achieving an accuracy of 97.50%, a macro F1-score of 98.51%, and a balanced accuracy of 98.18%. CatBoost and 1D-CNN also achieved highly competitive results, whereas the recurrent models produced substantially lower class-sensitive performance. SHAP analysis identified the sodium-to-potassium ratio, blood-pressure categories, and age as the most influential variables affecting the Random Forest predictions. The agreement between the exploratory analysis and the SHAP results supports the reliability of the selected feature relationships. Future work should evaluate the framework using larger and more diverse patient datasets, external validation cohorts, and repeated stratified evaluation to examine the stability and generalizability of the classification results.

REFERENCES

[1] M. Picard, M. Leclercq, A. Bodein, M. P. Scott-Boyer, O. Perin, and A. Droit, "Improving drug repositioning with negative data labeling using large language models," *Journal of Cheminformatics*, vol. 17, no. 1, p. 16, 2025.

[2] R. Tan, H. Hua, S. Zhou, Z. Yang, C. Yang, G. Huang, J. Zeng, and J. Zhao, "Current landscape of innovative drug development and regulatory support in china," *Signal Transduction and Targeted Therapy*, vol. 10, no. 1, p. 220, 2025.

[3] K. Zhang, X. Yang, Y. Wang, Y. Yu, N. Huang, G. Li, X. Li, J. C. Wu, and S. Yang, "Artificial intelligence in drug development," *Nature Medicine*, vol. 31, no. 1, pp. 45–59, 2025.

[4] A. Kumar, S. K. Gupta, and S. Kim, "AI-driven drug discovery using a context-aware hybrid model to optimize drug–target interactions," *Scientific Reports*, vol. 15, no. 1, p. 35719, 2025.

[5] R. Rodríguez-Pérez and J. Bajorath, "Evolution of support vector machine and regression modeling in chemoinformatics and drug discovery," *Journal of Computer-Aided Molecular Design*, vol. 36, no. 5, pp. 355–362, 2022.

[6] J. Kong, S. Park, T. H. Kim, J. E. Lee, H. Cho, J. Oh, S. Lee, H. Jo, H. Lee, K. Lee, J. Park, L. Jacob, D. Pizzol, S. Y. Rhee, S. Kim, and D. K. Yon, "Worldwide burden of antidiabetic drug-induced sarcopenia: An international pharmacovigilance study," *Archives of Gerontology and Geriatrics*, vol. 129, p. 105656, 2025.

[7] D. Li, H. Wang, C. Qin, D. Du, Y. Wang, Q. Du, and S. Liu, "Drug-induced acute pancreatitis: A real-world pharmacovigilance study using the FDA adverse event reporting system database," *Clinical Pharmacology & Therapeutics*, vol. 115, no. 3, pp. 535–544, 2024.

[8] R. Chelliah, M. Rubab, S. Vijayalakshmi, M. Karuvelan, K. Barathikannan, and D.-H. Oh, "Liposomes for drug delivery: Classification, therapeutic applications, and limitations," *Next Nanotechnology*, vol. 8, p. 100209, 2025.

[9] P. Uttreja, I. Karnik, A. Adel Ali Youssef, N. Narala, R. M. Elkanayati, S. Baisa, N. D. Alshammari, S. Banda, S. K. Vemula, and M. A. Repka, "Self-emulsifying drug delivery systems (SEDDS): Transition from liquid to solid—a comprehensive review of formulation, characterization, applications, and future trends," *Pharmaceutics*, vol. 17, no. 1, p. 63, 2025.

[10] A. Simitopoulos, A. Tsekouras, and P. Macheras, "Coupling drug dissolution with BCS," *Pharmaceutical Research*, vol. 41, no. 3, pp. 481–491, 2024.

[11] D. Benrimoh, C. Armstrong, J. Mehltretter, R. Fratila, K. Perlman, S. Israel, A. Kapelner, S. V. Parikh, J. F. Karp, K. Heller, and G. Turecki, "Development of the treatment prediction model in the artificial intelligence in depression–medication enhancement study," *npj Mental Health Research*, vol. 4, no. 1, p. 26, 2025.