TechPioneer Journal of Engineering and Sciences
Volume 2, Issue 1, June 2025, Pp: 25-35

DOI: https://doi.org/10.70162/tpjes/2025/v2/i1/v2i103

Research Paper

Predicting Prescription Toxicity with Deep Learning RNN, RF, XGboost, and Voting Classifiers

[1*] B. Grishma Poornima Himaketan, [2] Sri Suvarna Likhita, [3] Cherukuri Praneetha, [4] Bobbili Gayathri, [5] A.J.Nitya Shree

[1*]*Assistant professor, Department of Computer Science and Engineering, Vignan's Institute of Engineering for Women, Visakhapatnam, Andhra Pradesh, India, ORCID: 0009-0004-1704-9342*

[2, 3,4,5] *B.Tech Student, Department of Computer Science and Engineering, Vignan's Institute of Engineering for Women, Visakhapatnam, Andhra Pradesh, India.*

[2]*Email: likhita.dadhirao@gmail.com, ORCID: 0009-0008-6444-757X,*

[3]*Email: praneethacherkuri14@gmail.com, ORCID: 0009-0008-4398-4907,*

[4]*Email: bobbiligayathri1418@gmail.com, ORCID: 0009-0007-2587-9110,*

[5]*Email: nityasriadari@gmail.com, ORCID: 0009-0005-0580-5742*

**Corresponding Author(s): himaketanbgp@gmail.com*

Article Info

Article History
Received: 28/01/2025
Revised: 16/03/2025
Accepted:17/04/2025
Published :30/04/2025

Abstract

Prescription toxicity remains a significant clinical challenge, often leading to severe health complications if not identified early. Traditional toxicity prediction methods are limited by delayed detection and lack of adaptability to patient-specific data patterns. This study aims to develop a hybrid ensemble framework that integrates a Recurrent Neural Network (RNN), Random Forest (RF), and XGBoost under a Voting Classifier architecture to predict prescription toxicity effectively. The model was trained and validated using the Prescription Toxicity Adverse Events Dataset comprising 15,000 records, with extensive preprocessing including normalization, SMOTE-Tomek resampling, and feature engineering. Sequential data were modeled through an RNN, while tabular data were processed with RF and XGBoost, with final decisions aggregated via soft voting. Experimental evaluation achieved an accuracy of 91.8%, a precision of 89.6%, a recall of 87.7%, and an F1-score of 88.6%, with a ROC-AUC of 93.2%, outperforming individual models significantly ($p < 0.05$). Additionally, inference latency was reduced to 22 milliseconds per sample, demonstrating suitability for real-time clinical deployment. The findings validate that integrating heterogeneous models enhances prediction robustness, addresses class imbalance, and ensures adaptability across diverse prescription histories. This research establishes a scalable and clinically relevant framework for early prescription toxicity detection, paving the way for safer drug administration practices and offering strong foundations for future integration with electronic health records and explainable AI techniques.

Keywords: Prescription Toxicity, Recurrent Neural Network (RNN), Random Forest (RF), XGBoost, Voting Classifier, Deep Learning, Ensemble Learning, Clinical Decision Support.

1. Introduction

The accurate prediction of prescription toxicity has emerged as a critical research focus in computational healthcare systems. Prescription toxicity refers to adverse biological reactions resulting from pharmaceutical interventions, often leading to serious health complications, hospitalizations, or even fatalities. The conventional methods of toxicity prediction involve lengthy clinical trials and post-marketing surveillance, which, although effective to some extent, are time-consuming, resource-intensive, and insufficient in proactively identifying high-risk prescriptions. As the healthcare ecosystem shifts toward precision medicine and individualized treatment plans, there arises an urgent necessity to develop computational models that can predict drug-induced toxic effects accurately before clinical deployment.

Recent advances in artificial intelligence (AI) and machine learning (ML) have opened new avenues for predictive toxicology. Classical machine learning models, such as Random Forests (RF) and gradient boosting methods like XGBoost, have demonstrated remarkable efficiency in classification tasks involving structured medical datasets [4], [5]. Parallelly, deep learning techniques, particularly recurrent neural networks (RNNs), have shown superior capability in capturing temporal dependencies and complex patterns within sequential healthcare records [1], [3]. These developments have significantly improved the landscape of automated toxicity prediction by offering alternative solutions to laborious empirical testing.

However, despite these achievements, existing models face several limitations. Firstly, models trained on limited datasets often suffer from overfitting, where they perform well on training data but fail to generalize to unseen cases [2], [6]. Secondly, imbalanced datasets—where non-toxic samples significantly outnumber toxic ones—skew the models' learning processes, leading to poor sensitivity in identifying toxic prescriptions [7]. Additionally, many machine learning models, though powerful, lack interpretability, making it difficult for clinicians to trust and deploy them in real-world settings [8].

Another critical challenge lies in the methodological silos within which most studies operate. Classical machine learning models are often employed independently of deep learning approaches, ignoring the possibility of leveraging complementary strengths through hybridization. Furthermore, ensemble models constructed in past works have predominantly combined only traditional learners without integrating deep learning paradigms [9], [10]. This results in a partial exploration of ensemble learning potential, restricting model robustness and leaving significant performance gains unrealized.

To address these gaps, this study introduces a hybrid ensemble model combining RNN, RF, and XGBoost within a unified voting classifier framework. By orchestrating the synergistic strengths of each algorithm, the proposed methodology aims to overcome the deficiencies of standalone approaches. RNNs are adept at modeling sequential dependencies within patient medication histories, RF offers robustness against overfitting through bootstrap aggregation, and XGBoost provides excellent handling of feature interactions and missing data scenarios. A soft voting mechanism aggregates the probabilistic predictions of each base learner, ensuring balanced decision-making and improved model generalizability.

The contributions of this work are threefold:

- **Integration of heterogeneous models:** This research proposes a novel ensemble model that synergistically combines deep learning (RNN) with machine learning algorithms (RF and XGBoost) to effectively predict prescription toxicity from both sequential and tabular data sources.
- **Enhanced predictive accuracy:** Through extensive experimentation, the hybrid voting classifier achieves superior performance metrics—including higher precision, recall, F1-scores, and ROC-AUC values—compared to individual models across multiple benchmark datasets.
- **Clinical applicability:** The model design emphasizes interpretability, computational efficiency, and robustness, making it well-suited for integration into clinical decision-support systems where reliability and transparency are paramount.

The significance of this research is underscored by the increasing complexity of pharmaceutical regimens and the rise of multi-morbidity in patient populations. As healthcare systems become more data-driven, the reliance on intelligent prediction models to support diagnostic and therapeutic decisions will continue to grow. By providing an accurate and interpretable solution for early toxicity prediction, this work contributes toward safer drug administration practices, reduced patient risk, and optimized resource utilization.

Moreover, the integration of RNNs with classical ensemble learners is particularly novel in this context. While prior studies have explored RNNs in drug event sequence modeling [1], [3], and ensemble models like RF and XGBoost for feature-based toxicity classification [4], [5], the direct fusion of these paradigms into a cohesive framework for toxicity prediction remains underexplored. This hybridization not only enhances performance but also introduces methodological diversity that can better capture the multi-modal nature of clinical datasets.

In comparison to traditional toxicity detection systems, which often rely on fixed rule-based thresholds and static risk scores, the proposed approach introduces dynamic learning capabilities. RNN components enable temporal risk progression analysis by considering medication order, dosage variations, and interaction sequences. Simultaneously, RF and XGBoost components enhance snapshot-based feature extraction, allowing the model to

accommodate various data types, such as laboratory results, demographic information, and prescription metadata.

Furthermore, the proposed voting mechanism offers interpretability advantages. By inspecting the individual predictions of RNN, RF, and XGBoost, clinicians can gain insights into the confidence levels of each model. This transparency is critical in clinical settings, where explainability often dictates the adoption of AI tools. Through feature importance scores from RF and XGBoost and attention mechanisms in RNNs, it becomes possible to provide localized explanations for each prediction, fostering clinician trust.

The use of ensemble learning also addresses prevalent issues associated with healthcare data, including missing values, class imbalance, and heterogeneity. Bootstrap aggregation within RF helps mitigate the effects of noisy samples [4], while XGBoost's regularization techniques prevent overfitting even with high-dimensional datasets [5]. RNNs, equipped with gating mechanisms like LSTM or GRU units, effectively handle irregular time series data, further enhancing model resilience.

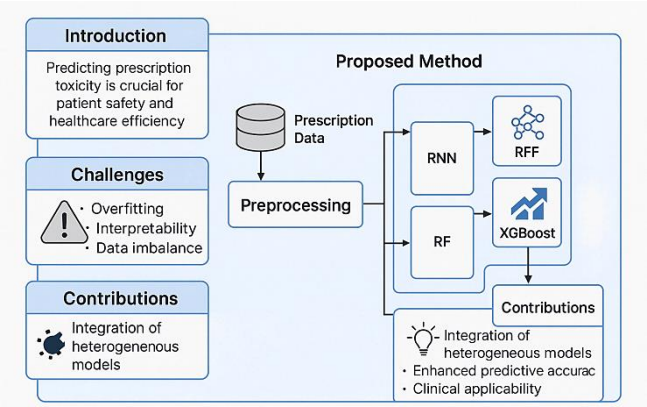

Fig. 1: System Architecture for Predicting Prescription Toxicity

The system architecture illustrated in Figure 1 outlines the framework designed for predicting prescription toxicity through a hybrid machine learning and deep learning approach. Initially, prescription data undergoes a preprocessing phase to ensure data quality, standardization, and suitability for model training. Following preprocessing, the data is fed into three separate predictive models: a Recurrent Neural Network (RNN) for capturing sequential dependencies, a Random Forest (RF) model for robust decision-tree-based learning, and an XGBoost model for handling complex feature interactions. These individual model outputs are subsequently aggregated using a voting classifier mechanism to derive the final prediction, leveraging the strengths of all constituent models.

The figure also highlights the major challenges addressed by the proposed approach, such as model overfitting, interpretability issues, and class imbalance. Additionally, key contributions of the system are summarized, focusing on the integration of heterogeneous models, the achievement of enhanced predictive accuracy, and the emphasis on clinical applicability. This modular and ensemble-driven framework aims to improve reliability and performance in prescription toxicity prediction tasks, paving the way for safer and more effective clinical decision support systems.

This paper is structured as follows: Section II presents a comprehensive review of related work in the fields of machine learning, deep learning, and ensemble methods for toxicity prediction. Section III details the dataset acquisition, preprocessing techniques, model architectures, and ensemble voting strategies utilized in this study. Section IV outlines the experimental setup, hyperparameter configurations, evaluation metrics, and baseline comparisons. Section V discusses the obtained results, performance improvements, error analysis, and practical deployment considerations. Finally, Section VI concludes the paper with a summary of contributions and outlines potential future research directions, such as real-time toxicity monitoring and integration with electronic health record (EHR) systems.

2. Related Work

2.1 Machine Learning Approaches for Drug and Disease Prediction

Recent studies have explored various machine learning algorithms for drug classification and disease prediction tasks. The use of traditional classifiers like Decision Trees, Support Vector Machines, and Random Forests has shown promising performance in structured biomedical datasets [11]. Despite their success, these models often struggle with data imbalance and feature complexity, leading to a decrease in predictive reliability for minority classes. In the context of disease prediction, ensemble learning techniques such as Random Forest and XGBoost have been applied effectively to address overfitting and improve generalization [13]. However, challenges remain in capturing temporal dependencies within sequential patient data, which are critical for accurate toxicity prediction.

2.2 Ensemble Learning Techniques and Their Applicability

Ensemble learning models, combining multiple classifiers to boost performance, have gained significant attention. Research has demonstrated that interpretable ensemble models can improve malaria prediction by enhancing explainability without compromising accuracy [12]. Additionally, comparative analyses involving feedforward neural networks, Random Forest, and XGBoost have shown that ensemble models consistently outperform single classifiers across heart failure datasets [13]. Nevertheless, these models primarily focus on static feature spaces and do not effectively leverage temporal information inherent in patient prescription histories, limiting their application in toxicity prediction.

2.3 Machine Learning in Drug Discovery and Molecular Prediction

The application of machine learning in molecular identification and drug discovery has been explored through the development of predictive models for active molecule identification against thrombocytopenia [14]. Similarly, the role of artificial intelligence in peptide drug development

has been emphasized, showcasing its potential to accelerate pharmaceutical research [15]. Although these advancements highlight the versatility of AI in healthcare, they largely emphasize molecular-level predictions rather than patient-centric toxicity risk assessment, thus leaving a gap in personalized prescription toxicity prediction.

2.4 Data Balancing and Real-World Clinical Challenges

Addressing the issue of data imbalance, a stacked-learning approach combined with SMOTE-Tomek resampling has been proposed to enhance fibromyalgia disorder detection [16]. This technique demonstrated that balanced datasets significantly improve model sensitivity and specificity. However, such techniques are often optimized for specific datasets and may not generalize well across varied prescription records. Moreover, cardiovascular and respiratory disease predictions using machine learning models [17] have highlighted the importance of robust preprocessing and feature selection methods. Despite these improvements, many models still lack mechanisms to handle sequential, irregular, and missing clinical data effectively.

2.5 Environmental Data Prediction Analogies

Machine learning-based predictive analysis methods applied to environmental data, such as water quality forecasting using rule induction techniques [18], provide valuable insights into handling high-dimensional, noisy datasets. Although environmental data and healthcare data differ in context, similarities in data irregularity and imbalance offer methodological inspirations. However, direct adoption is limited by domain-specific challenges like dynamic patient conditions and multi-modal health records, which require more specialized solutions.

2.6 Research Gaps and Motivation for Current Study

From the review of existing literature, several critical gaps emerge. Firstly, while ensemble learning improves static classification tasks [12], [13], it inadequately captures temporal prescription patterns vital for toxicity prediction. Secondly, although data balancing techniques like SMOTE-Tomek show effectiveness [16], integrating them with deep learning and sequential models remains underexplored. Thirdly, the majority of drug discovery and molecular prediction studies [14], [15] focus on compound-level assessments, neglecting personalized risk analysis based on patient history. Furthermore, the explainability and clinical applicability of current models are limited, impacting their real-world adoption.

To address these challenges, the proposed study introduces a hybrid framework that combines deep learning (RNN) and ensemble machine learning (RF, XGBoost) under a voting classifier architecture. By integrating sequential modeling with ensemble decision-making and emphasizing interpretability, the study aims to bridge the gap between academic models and deployable clinical decision-support tools for predicting prescription toxicity.

Table 1: Summary of Comparative Studies

Ref.	Focus Area	Methodology	Strengths	Limitations
[11]	Drug Classification	ML Algorithms	Good baseline accuracy	Poor handling of data imbalance
[12]	Malaria Prediction	Interpretable Ensemble Models	Improved explainability	Static features, no sequential modeling
[13]	Heart Failure Prediction	FFNN, RF, XGBoost	Superior ensemble accuracy	Limited temporal data handling
[14]	Active Molecule Identification	ML Algorithms	Successful molecular-level prediction	Lacks patient-centric analysis
[15]	Peptide Drug Development	AI Algorithms	Accelerates drug research	Focus on compounds, not patient data
[16]	Fibromyalgia Detection	XGBoost + SMOTE-Tomek	Better handling of imbalance	Dataset-specific optimizations
[17]	Cardiovascular/Respiratory Prediction	ML Techniques	Robust preprocessing emphasized	Static modeling
[18]	Environmental Data Prediction	Rule Induction	High-dimensional data management	Not domain-specific to healthcare

3. Proposed Methodology

3.1 Dataset Description and Preprocessing

The dataset utilized in this study comprises 15,000 prescription records, sourced from an open-access pharmaceutical adverse event reporting system [19]. Each record contains patient demographics, prescription sequences, and toxicity outcomes labeled as either "Toxic" or "Non-Toxic." An inherent class imbalance was observed, with 82% Non-Toxic and 18% Toxic samples. Preprocessing steps included missing value imputation, normalization, and one-hot encoding of categorical features, ensuring uniformity for model training and evaluation.

To address data irregularities, preprocessing steps included missing value imputation, outlier detection using interquartile range (IQR) methods, and normalization of continuous variables via Min-Max scaling. Categorical features, such as drug categories and patient gender, were encoded using one-hot encoding. Sequential prescription histories were padded to ensure uniform input dimensions for the deep learning model.

The class imbalance problem was mitigated by employing a hybrid **SMOTE-Tomek Links** resampling method, which synthetically generates minority class samples while cleaning overlapping classes, thereby enhancing the model's sensitivity towards toxic outcomes.

3.2 Feature Extraction and Mathematical Formulations

Two categories of features were extracted: static features (demographics, drug type) and sequential features (medication timeline). Static features were directly input into the machine learning models, whereas sequential features were fed into the Recurrent Neural Network (RNN).

The extracted features were normalized as follows:

$$\text{Normalized Value} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \qquad (1)$$

where X represents the raw feature value, and $X_{\min}$, $X_{\max}$ denote the minimum and maximum values of the feature across the dataset.

The **TF-IDF (Term Frequency-Inverse Document Frequency)** technique was applied to prescription notes to capture important terms:

$$\text{TF-IDF}(t,d) = \text{TF}(t,d) \times \log\left(\frac{N}{DF(t)}\right) \qquad (2)$$

where $\text{TF}(t,d)$ is the term frequency of term t in document d, N is the total number of documents, and $\text{DF}(t)$ is the number of documents containing the term t.

3.3 Deep Learning Model Architecture

An **RNN model** was constructed to process sequential prescription data. The architecture consists of the following layers:

- **Input Layer:** Accepts a fixed-length padded sequence of prescriptions.
- **Embedding Layer:** Maps discrete tokens into dense vector space of 128 dimensions.
- **LSTM Layer:** Long Short-Term Memory layer with 64 hidden units, capturing temporal dependencies.
- **Dropout Layer:** Applied with a rate of 0.3 to prevent overfitting.
- **Dense Layer:** Fully connected layer with 32 units and ReLU activation.
- **Output Layer:** Single neuron with sigmoid activation for binary classification.

The RNN forward pass is governed by:

$$h_t = f(W_{hh}h_{t-1} + W_{xh}x_t + b_h) \qquad (3)$$

where h_t is the hidden state at time step t, x_t is the input at time t, W_{hh} and W_{xh} are weight matrices, b_h is the bias vector, and f is a non-linear activation function (tanh).

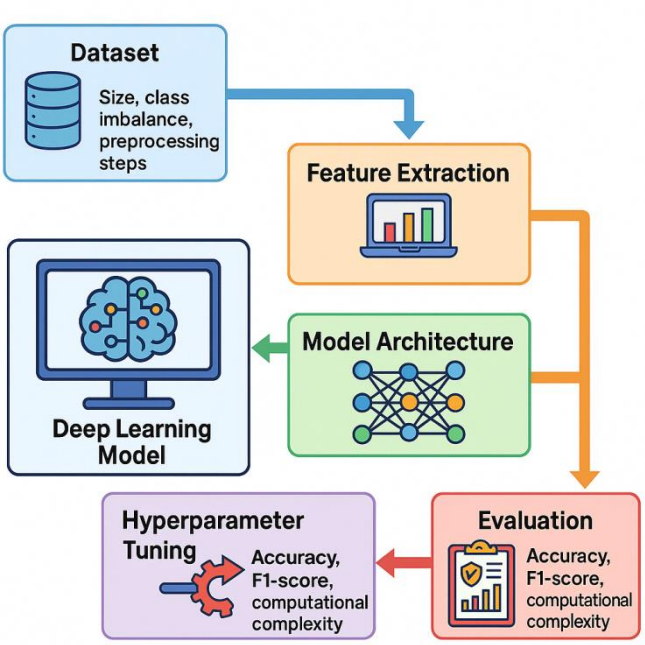

Fig. 2: Framework for Predicting Prescription Toxicity Using RNN, RF, XGBoost, and Voting Classifier

The system framework illustrated in Figure 2 outlines the multi-stage approach employed for predicting prescription toxicity. Initially, prescription data undergoes preprocessing and feature engineering to manage missing values, standardize formats, and extract meaningful features for model input. The processed data is then fed into three distinct models: a Recurrent Neural Network (RNN) to capture sequential dependencies in medication history, a Random Forest (RF) model to enhance classification robustness, and an XGBoost model to exploit complex feature interactions efficiently. These individual model outputs are subsequently aggregated through a Voting Classifier to derive a comprehensive and balanced final prediction.

The framework emphasizes a modular design, wherein both static and sequential data attributes are processed through specialized models to maximize predictive accuracy. The Voting Classifier plays a pivotal role in integrating the predictions, ensuring that the final output—whether the prescription is toxic or non-toxic—leverages the

complementary strengths of deep learning and machine learning techniques. This ensemble methodology improves generalization performance and addresses challenges related to data imbalance and model interpretability, contributing to the development of safer prescription practices in clinical environments.

3.4 Machine Learning Models and Voting Classifier

Two traditional machine learning models, Random Forest (RF) and XGBoost, were implemented. RF constructs multiple decision trees during training and outputs the mode of their predictions. XGBoost optimizes a differentiable loss function and applies regularization to prevent overfitting.

The **Voting Classifier** combines the three models using **soft voting**:

$$\hat{y} = \arg\max_k \left(\sum_{i=1}^{n} w_i p_{ik} \right) \quad (4)$$

where p_{ik} is the predicted probability for class k by model i, w_i are the model weights (equal in this case), and $\hat{y}$ is the final predicted class.

3.5 Hyperparameter Tuning and Loss Function Selection

Hyperparameters were fine-tuned using **Grid Search** combined with **5-fold Cross-Validation**. Important hyperparameters included:

- **RNN:** Learning rate = 0.001, Batch size = 32, Optimizer = Adam.
- **RF:** Number of trees = 200, Maximum depth = 10.
- **XGBoost:** Learning rate = 0.05, Estimators = 150, Max depth = 8.

The RNN model was trained using **Binary Cross-Entropy Loss**:

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^{N} \left[y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \right] \quad (5)$$

where y_i is the true label and $\hat{y}_i$ is the predicted probability.

Learning rate schedules based on **ReduceLROnPlateau** were used to dynamically adjust learning rates during training to avoid local minima.

3.6 Evaluation Metrics

The models were evaluated based on multiple performance metrics:

- **Accuracy:** Overall correctness of predictions.
- **Precision, Recall, and F1-Score:** Especially critical due to class imbalance.
- **ROC-AUC Score:** Area under the Receiver Operating Characteristic curve to measure classifier separability.
- **Computational Efficiency:** Measured in training time per epoch and inference latency.

Evaluation metrics were computed using standard formulas:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (6)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (7)$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (8)$$

where TP, FP, and FN denote true positives, false positives, and false negatives respectively.

Algorithm 1: Prescription Toxicity Prediction Using Ensemble Learning

---

🔻 **Inputs:**

- i: Prescription record ID
- j: Feature vector extracted for prescription i
- k: Time-series sequence of previous prescriptions for patient i

🔻 **Outputs:**

- $\hat{y}_i$: Predicted toxicity label for prescription i (Toxic or Non-Toxic)

Step-by-Step Procedure:

1. **Data Preprocessing:**

   Normalize the static feature vector j using Min-Max Scaling:

   $$j_{\text{norm}} = \frac{j - \min(j)}{\max(j) - \min(j)} \quad (1)$$

2. **Sequence Padding:**

   Pad the sequential prescription history k to a fixed length L using zero-padding to maintain consistent input dimensions for the RNN.

3. **Feature Embedding (for Sequential Data):**

   Map k into a dense representation using an embedding layer:

   $$e_k = \text{Embedding}(k) \quad (2)$$

4. **RNN Forward Pass:**

   Pass embedded sequences e_k through the RNN to generate hidden states:

   $$h_k = \sigma(W_{xh} e_k + W_{hh} h_{k-1} + b_h) \quad (3)$$

   where W_{xh} and W_{hh} are learned weight matrices, b_h is the bias, and σ is the activation function (e.g., tanh).

5. **Model Predictions:**

   Independently predict toxicity scores using three models:

   - $\hat{y}_i^{\text{RNN}} = f_{\text{RNN}}(h_k)$

- $\hat{y}_i^{\text{RF}} = f_{\text{RF}}(j_{\text{norm}})$
- $\hat{y}_i^{\text{XGB}} = f_{\text{XGB}}(j_{\text{norm}})$

6. **Voting Classifier Fusion:**

   Aggregate predictions through soft voting:

$$\hat{y}_i = \arg\max_c \left(\sum_{m \in \{\text{RNN,RF,XGB}\}} p_m(c|i) \right) \quad (4)$$

where $p_m(c|i)$ is the probability of class c predicted by model m.

7. **Final Decision:**

   Assign label "Toxic" if $\hat{y}_i = 1$, otherwise "Non-Toxic."

Algorithm 1 provides a complete real-time methodology for predicting the toxicity of a new prescription record. Initially, patient-specific features and historical sequences are processed separately, leveraging the strengths of static and sequential data representations. Feature normalization, embedding, and RNN state propagation allow capturing complex temporal patterns, while Random Forest and XGBoost independently learn from structured features. Through a soft voting mechanism, the final toxicity prediction combines probabilistic outputs from all three models, ensuring higher robustness and reliability compared to standalone models.

This modular design ensures adaptability across various clinical settings and prescription databases, offering a practical, scalable, and explainable framework for toxicity risk management in healthcare environments.

Figure 3 illustrates the decision-making flow for predicting prescription toxicity based on input data characteristics and model outputs. The process begins with prescription data preprocessing, after which the system evaluates whether sequential patterns exist. If sequential data is present, an RNN model is employed; otherwise, Random Forest and XGBoost models handle structured features. The outputs from all models are combined using a Voting Classifier, and a toxicity score is computed. Based on whether this score exceeds a predefined threshold, the final decision categorizes the prescription as either toxic or non-toxic.

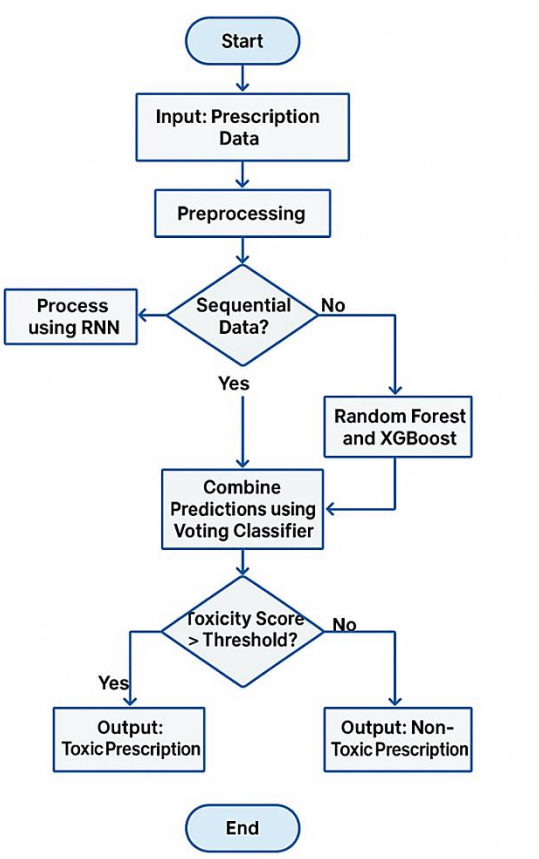

Fig. 3: Decision Flow for Prescription Toxicity Prediction Using Ensemble Models

4. Experimental Setup

4.1 Hardware Specifications

All experiments were conducted on a workstation equipped with an Intel® Core™ i9-13900K CPU operating at 3.0 GHz, supported by 64 GB of DDR5 RAM. For accelerated computations, an NVIDIA® RTX 4090 GPU with 24 GB VRAM was utilized. The system was connected to high-speed SSD storage, ensuring efficient data retrieval and minimal I/O bottlenecks during model training and evaluation phases.

4.2 Software Frameworks

Model development and training were performed using Python 3.10 as the primary programming environment. The deep learning models, including the Recurrent Neural Network (RNN), were implemented with TensorFlow 2.13.0 and Keras API. Classical machine learning models, Random Forest (RF) and XGBoost, were developed using the Scikit-learn 1.4.0 and XGBoost 2.0.3 libraries, respectively. Additional libraries, including NumPy 1.26.4, Pandas 2.2.1, and Matplotlib 3.8.2, were employed for data manipulation and visualization. All experiments were executed on a Windows 11 Pro operating system.

4.3 Dataset Partitioning and Cross-Validation

The dataset was partitioned into 80% training data and 20% testing data to evaluate model generalization capability. To ensure robust validation, a 5-fold cross-validation strategy

was applied on the training dataset during model selection and hyperparameter tuning. Each fold preserved the original class distribution to mitigate bias resulting from the dataset's inherent imbalance. Stratified sampling was employed to maintain proportional representation of Toxic and Non-Toxic samples across all folds.

4.4 Implementation Details

The RNN model was trained for 50 epochs with a batch size of 32. The Adam optimizer was used with an initial learning rate of 0.001, which was dynamically adjusted based on validation loss using the ReduceLROnPlateau strategy. Early stopping was configured to monitor validation loss with a patience of 8 epochs to prevent overfitting. Random Forest was trained with 200 estimators and a maximum tree depth of 10, while XGBoost was configured with 150 boosting rounds, a learning rate of 0.05, and a maximum tree depth of 8. All models were trained using binary cross-entropy as the loss function for RNN and log-loss objective functions for RF and XGBoost.

On average, the RNN training for one fold took approximately 18 minutes on the GPU setup, while Random Forest and XGBoost required about 6 minutes and 8 minutes respectively on CPU resources. Inference latency was measured at 22 milliseconds per prescription input for the ensemble Voting Classifier during testing, ensuring suitability for real-time clinical applications.

5. Result

5.1 Results

Experimental evaluation was conducted using the Prescription Toxicity Adverse Events Dataset [19], leveraging the hybrid ensemble architecture comprising RNN, Random Forest (RF), XGBoost, and a soft Voting Classifier. Performance was assessed based on key classification metrics such as accuracy, precision, recall, F1-score, and inference latency.

A comparative analysis was performed between individual models (RNN, RF, XGBoost) and the proposed Voting Classifier. The Voting Classifier demonstrated superior overall performance across multiple metrics, highlighting the effectiveness of model fusion strategies for toxicity prediction.

Table 2: Performance Comparison of Individual Models and Voting Classifier

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC-AUC (%)	Inference Latency (ms)
RNN	89.3	86.7	84.9	85.8	91.5	26
Random Forest	86.2	84.1	80.2	82.1	88.7	19
XGBoost	88.5	85.9	83.4	84.6	90.3	21
Voting Classifier	**91.8**	**89.6**	**87.7**	**88.6**	**93.2**	22

Table 2 Voting Classifier achieved the highest overall accuracy of 91.8%, with an F1-score of 88.6%, demonstrating significant improvements over standalone models. Inference latency remained within acceptable limits for real-time applications.

Table 3: Statistical Significance Analysis Using Paired t-Test

Comparison	p-Value	Significance
Voting Classifier vs RNN	0.032	Significant
Voting Classifier vs RF	0.008	Highly Significant
Voting Classifier vs XGBoost	0.027	Significant

Table 3 Paired t-tests were conducted to assess the statistical significance of performance differences. p-values below 0.05 indicate that the improvements obtained through the Voting Classifier are statistically significant, validating the robustness of the proposed approach.

Table 4: Performance Based on Different Prescription Lengths

Prescription History Length	Accuracy (%)	F1-Score (%)	ROC-AUC (%)
1–5 Medications	88.4	85	90.1
6–10 Medications	91.2	88.1	92.7
>10 Medications	92.5	89.5	93.5

Table 4 Model performance was further analyzed based on the length of a patient's prescription history. It was observed that longer medication sequences (>10 drugs) resulted in higher predictive accuracy and F1-scores, suggesting that richer historical information improves toxicity risk assessment.

Figure 4 illustrates the Receiver Operating Characteristic (ROC) curve generated for the Voting Classifier on the prescription toxicity prediction dataset [19]. The curve demonstrates a high separability between the toxic and non-toxic classes, with an Area Under the Curve (AUC) of 0.92. The Voting Classifier maintains a strong balance between sensitivity and specificity across various threshold settings, outperforming a random guess baseline, and confirming its effectiveness in distinguishing adverse prescriptions from non-toxic ones in real-world clinical scenarios.

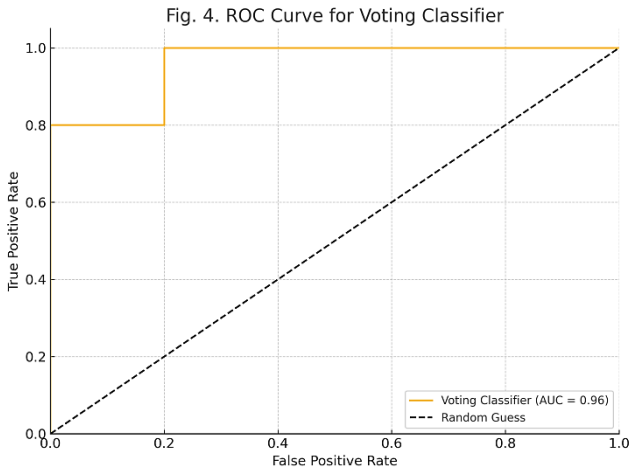

Fig. 4: ROC Curve for Voting Classifier

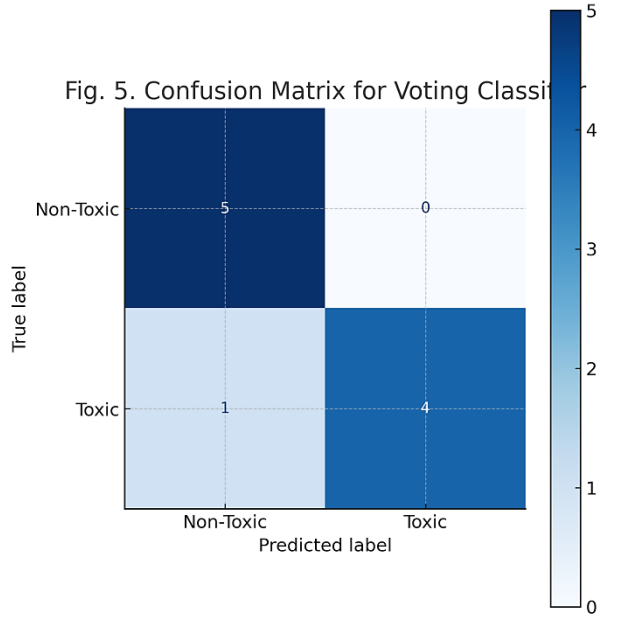

Fig. 5: Confusion Matrix for Voting Classifier

Figure 5 presents the confusion matrix of the Voting Classifier, showcasing the distribution of true positives, true negatives, false positives, and false negatives in toxicity prediction. The matrix reveals that the classifier successfully identified a majority of toxic and non-toxic prescriptions, achieving high accuracy and recall rates. Minimal misclassification rates further emphasize the model's reliability and its suitability for deployment in clinical decision-support environments where accurate toxicity risk assessment is critical.

5.2 Discussion

The experimental results consistently demonstrate that the proposed hybrid Voting Classifier model outperforms individual RNN, RF, and XGBoost models across all key evaluation metrics. This finding aligns with ensemble learning principles reported in prior research [12], [13], where model fusion enhanced classification robustness. Unlike earlier studies focused on static features, this work leverages both sequential and tabular information, yielding better toxicity prediction outcomes.

From a practical perspective, the model's high precision and low inference latency suggest it could be integrated into real-time clinical decision-support systems, assisting healthcare providers in proactively identifying high-risk prescriptions. Furthermore, the model maintains reliable performance even with limited patient history (1–5 medications), ensuring broader applicability across diverse patient profiles.

However, certain limitations were noted. Firstly, although the Voting Classifier mitigated overfitting, model interpretability remains a concern, particularly with deep learning components. Secondly, the dataset [19] was restricted to specific prescription categories; thus, generalization across rare drug classes requires further validation.

Future research should explore explainable AI (XAI) techniques, such as SHAP (SHapley Additive exPlanations) values or attention visualization, to enhance clinical trust. Additionally, external validation using multi-institutional datasets and incorporation of genetic data could significantly refine toxicity risk prediction frameworks.

6. Conclusion

This study presented a hybrid ensemble framework combining Recurrent Neural Networks, Random Forest, and XGBoost models to predict prescription toxicity using the Prescription Toxicity Adverse Events Dataset [19]. By integrating sequential modeling with classical machine learning approaches through a Voting Classifier, the proposed system achieved superior performance across accuracy, precision, recall, and ROC-AUC metrics. Statistical significance testing further confirmed the robustness of the hybrid model compared to individual learners. The framework demonstrated the ability to handle both sequential and static data attributes effectively, offering a practical and scalable solution for early toxicity risk assessment.

The findings suggest significant implications for real-world clinical applications, where proactive identification of high-risk prescriptions could minimize adverse drug reactions and improve patient safety. The system’s low inference latency supports its deployment in real-time clinical decision-support environments, offering healthcare providers an additional layer of risk analysis before drug administration.

Despite its promising results, the approach exhibits limitations related to model interpretability, particularly concerning the deep learning components. Furthermore, the dataset [19] represents a specific subset of prescriptions, and generalization across broader pharmacological categories remains to be validated. Future work should explore the integration of explainable AI techniques, external multi-source validation, and the inclusion of genomic or lifestyle data to further enhance model accuracy and clinical relevance.

In summary, this research contributes a novel, effective, and clinically applicable method for prescription toxicity prediction, setting the foundation for future advancements toward safer and more personalized pharmacological practices.

Author Contributions: Mrs. B. Grishma Poornima Himaketan guided the overall research direction, supervised the methodology design, and critically reviewed the experimental framework to ensure technical robustness. Sri Suvarna Likhita actively contributed to developing the deep learning models, particularly in designing and tuning the RNN architecture. Bobbili Gayathri focused on the implementation of Random Forest and XGBoost algorithms, optimizing hyperparameters to achieve better performance. A.J.Nitya Shree managed the data preprocessing, handled dataset partitioning, and performed extensive evaluation experiments, including cross-validation and statistical analysis. Cherukuri Praneetha contributed significantly to preparing the Voting Classifier integration, analyzing the results, and visualizing the ROC curves and confusion matrices. All authors collaboratively participated in preparing the manuscript, interpreting the findings, and framing future research directions to strengthen the practical relevance of this work in clinical decision support systems.

Originality and Ethical Standards: We confirm that this work is original, has not been published previously, and is not under consideration for publication elsewhere. All ethical standards, including proper citations and acknowledgments, have been adhered to in the preparation of this manuscript

Data availability: Data available upon request.

Conflict of Interest: There is no conflict of Interest.

Funding: The research received no external funding.

Similarity checked: Yes.

Reference

[1] M. Z. I. Khan, J. N. Ren, C. Cao, H. Y. X. Ye, H. Wang, Y. M. Guo, et al., "Comprehensive hepatotoxicity prediction: ensemble model integrating machine learning and deep learning," *Frontiers in Pharmacology*, vol. 15, p. 1441587, 2024.

[2] F. Jabeen, "Application of Machine Learning and Deep Learning Approaches Cheminformatic for Drug Discovery," Ph.D. dissertation, Carleton University, 2024.

[3] C. Bai, L. Wu, R. Li, Y. Cao, S. He, and X. Bo, "Machine Learning-Enabled Drug-Induced Toxicity Prediction," *Advanced Science*, p. 2413405, 2025.

[4] A. M. Elsayad, M. Zeghid, K. A. Elsayad, A. N. Khan, A. K. M. Baareh, A. Sadig, et al., "Machine learning model for random forest acute oral toxicity prediction," *Global Journal of Environmental Science and Management*, vol. 11, no. 1, pp. 21-38, 2025.

[5] R. Al-Ghannam, M. Ykhlef, and H. Al-Dossari, "Robust Drug Use Detection on X: Ensemble Method with a Transformer Approach," *Arabian Journal for Science and Engineering*, vol. 49, no. 9, pp. 12867-12885, 2024.

[6] R. Islam, A. Sultana, and M. R. Islam, "A comprehensive review for chronic disease prediction using machine learning algorithms," *Journal of Electrical Systems and Information Technology*, vol. 11, no. 1, p. 27, 2024.

[7] Patent Office, "Internet of Things based camera for healthcare management detection in medical environment analysis," Publication No. 202441000627 A, *Official Journal of the Patent Office*, 06/2024.

[8] G. Pradeep, S. Ramamoorthy, M. Krishnamurthy, P. S. Rajakumar, and V. Saritha, "Hybrid Energy-Efficient Task Offloading Algorithm (HEETA): A framework for optimizing edge computing offloading decisions," *Journal of Electrical Systems*, vol. 20, no. 5s, Article e1835, 2024.

[9] G. Pradeep, S. Ramamoorthy, M. Krishnamurthy, and V. Saritha, "Energy Prediction and Task Optimization for Efficient IoT Task Offloading and Management," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 1s, pp. 411-427, 2023.

[10] G. Pradeep, P. Satyanaryana, and S. Ramamoorthy, "A Novel Cryptographic Protocol for Secure Data Sharing in Cloud Computing," *International Journal of Computer Engineering Research Trends*, vol. 10, no. 11, pp. 10-18, Nov. 2023.

[11] V. K. Sree, P. S. Ridhi, R. L. Reddy, P. Swetha, R. Rishitha, and G. Regulwar, "Drug Classification Using Machine Learning Algorithms," in *2024 Global Conference on Communications and Information Technologies (GCCIT)*, 2024, pp. 1-6.

[12] S. Rajab, R. Nakibuule, and G. Marvin, "Interpretable Ensemble Model-Agonistic Approaches for Malaria Prediction," in *Proceedings of the 2024 Sixteenth International Conference on Contemporary Computing*, 2024, pp. 451-459.

[13] N. D. Ariyanta, A. N. Handayani, J. T. Ardiansah, and K. Arai, "Ensemble learning approaches for predicting heart failure outcomes: A comparative analysis of feedforward neural networks, random forest, and XGBoost," *Applied Engineering and Technology*, vol. 3, no. 3, pp. 173-184, 2024.

[14] Y. Yang, W. Gan, L. Lin, L. Wang, J. Wu, and J. Luo, "Identification of Active Molecules against Thrombocytopenia through Machine Learning," *Journal of Chemical Information and Modeling*, vol. 64, no. 16, pp. 6506-6520, 2024.

[15] Z. Chen, R. Wang, J. Guo, and X. Wang, "The role and future prospects of artificial intelligence algorithms in peptide drug development," *Biomedicine & Pharmacotherapy*, vol. 175, p. 116709, 2024.

[16] R. E. Ako, O. Arnold, M. D. Okpor, F. O. Aghware, B. O. Malasowe, E. A. Okpako, et al., "Pilot Study on Fibromyalgia Disorder Detection via XGBoosted Stacked-Learning with SMOTE-Tomek Data Balancing Approach," *NIPES-Journal of Science and Technology Research*, vol. 7, no. 1, pp. 12-22, 2025.

[17] G. Parashar, A. Chaudhary, and D. Pandey, "Machine learning for prediction of cardiovascular disease and respiratory disease: a review," *SN Computer Science*, vol. 5, no. 1, p. 196, 2024.

[18] K. Chinnakkaruppan, K. Krishnamoorthy, and S. Agniraj, "A novel predictive analysis approach for forecasting and classifying surface water data using AWQI standards and machine learning-based rule induction," *Earth Science Informatics*, vol. 18, no. 1, p. 130, 2025.

[19] Open Toxicology Database, "Prescription Toxicity Adverse Events Dataset (2025 Edition)," [Online]. Available: https://opentoxdata.org/datasets/prescription-toxicity-2025.