GSC Biological and Pharmaceutical Sciences
eISSN: 2581-3250 CODEN (USA): GBPSC2
Cross Ref DOI: 10.30574/gscbps
Journal homepage: https://gscbps.gsconlinepress.com/

(RESEARCH ARTICLE)

EXPLAINABLE ARTIFICIAL NEURAL NETWORKS FOR TABULAR DATA: IDENTIFYING STRUCTURAL, ELECTRONIC AND PHYSICOCHEMICAL DETERMINANTS OF PHARMACEUTICAL COMPOUND-INDUCED RESPIRATORY TOXICITY

Désiré Mélèdge, Aubin N'guessan *, Kouakou Jean Louis Kouakou and Donatien Zamblé

Fundamental Applied Physics Laboratory (FAPL), Nangui Abrogoua University, Côte d'Ivoire.
* Corresponding Author

ORCID Details
Désiré Mélèdje: https://orcid.org/0009-0008-0089-6595
Aubin N'guessan: https://orcid.org/0000-0002-3043-2985
Kouakou Jean Louis Kouakou https://orcid.org/0009-0006-3725-433X

GSC Biological and Pharmaceutical Sciences, 2026, 37(02), 001–008

Publication history: Received on 18 September 2026; revised on 27 October 2026; accepted on 29 October 2026

Article DOI: https://doi.org/10.30574/gscbps.2026.37.2.0372

Copyright © 2026Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution License 4.0

ABSTRACT

Respiratory toxicity induced by pharmaceutical compounds is a major concern in drug discovery and development, emphasizing the need for reliable computational approaches to identify potentially hazardous compounds at an early stage. In this study, an artificial neural network (ANN) integrated with eXplainable Artificial Intelligence (XAI) was developed for the binary classification of pharmaceutical compounds as respiratory toxic or non-toxic using a tabular dataset of molecular descriptors. Model performance was evaluated through ten-fold cross-validation and external validation using accuracy, precision, recall, F1-score, Matthews correlation coefficient (MCC), and area under the receiver operating characteristic curve (AUC-ROC). During internal validation, the ANN achieved an accuracy of 84.05% and an AUC of 91.36%. On the external test set, the model yielded an accuracy of 83.40%, precision of 85.58%, recall of 85.19%, F1-score of 85.38%, MCC of 0.66, and AUC-ROC of 89.53%, demonstrating robust predictive performance and satisfactory generalization to unseen compounds. SHapley Additive exPlanations (SHAP) analysis was further employed to interpret the ANN predictions and identify the molecular characteristics most strongly associated with model output. The most influential descriptors included nBase, BCUTp-1h, SssO, minssCH2, JGI8, maxHBd, nHeteroRing, GATS2m, ETA_BetaP_ns_d, GATS1e. Collectively, these descriptors encode complementary information related to basicity, heterocyclic structure, local electronic environments, hydrogen-bonding characteristics, atomic polarizability, charge distribution, atomic mass, electronegativity, and molecular topology. The SHAP findings suggest that respiratory toxicity predictions are governed by the combined contribution of multiple structural and physicochemical characteristics rather than by a single molecular property.

Keywords: Respiratory toxicity; Pharmaceutical compounds; Artificial neural networks; Molecular descriptors; SHAP; Explainable artificial intelligence.

1 INTRODUCTION

Among the various organs and physiological systems susceptible to the toxic effects of pharmaceutical compounds, the respiratory system represents a major target. Its vulnerability is particularly significant because the respiratory tract also serves as an important route of exposure and absorption for inhaled gases, airborne particles, and other potentially toxic substances. Respiratory toxicity can lead to a wide range of adverse health effects, including pneumonia, bronchitis, asthma, rhinitis etc. Because drug-induced respiratory toxicity may remain undetected during the early stages of drug development, assessing the respiratory toxicity risk of pharmaceutical compounds as early as possible is crucial for improving drug safety and reducing late-stage failures [1]. Computational approaches, particularly quantitative structure–toxicity relationship (QSTR) modeling, can leverage large publicly available databases containing thousands of compounds with toxicity endpoints expressed as binary classes, multiple classes, or continuous values to develop predictive models based on topological, electronic and physicochemical molecular descriptors. However, conventional QSTR approaches relying on simple regression models may be limited in their ability to capture the complex, nonlinear relationships and underlying patterns within high-dimensional molecular data. Consequently, increasing research efforts have focused on integrating artificial intelligence (AI) and machine learning algorithms into QSTR modeling to improve the prediction of molecular toxicity [2].

To date, several QSTR models have been developed to classify compounds according to their respiratory toxicity. N'Guessan et al. proposed an efficient Random Forest (RF) model incorporating a data-mining-based feature selection approach. From an initial set of 1,544 molecular descriptors, only nine relevant descriptors were retained, resulting in a parsimonious model with good predictive performance and an AUC of 0.83 on the test set [3]. Since tree-based methods are relatively insensitive to data heterogeneity and incorporate randomization mechanisms through both bootstrap sampling and feature selection, Jaganathan and coworkers developed an explainable artificial intelligence (XAI) model that achieved an accuracy of 86.9% and a Matthews correlation coefficient (MCC) of 0.734 [4]. Jaganathan et al. also employed a Support Vector Machine (SVM)-based machine learning approach to develop a binary classification model for predicting respiratory toxicity, providing a useful computational tool for the early identification and assessment of potentially respiratory-toxic compounds during drug development [5].

To the best of our knowledge, relatively few predictive models based on artificial neural networks (ANN) have been developed for assessing the respiratory toxicity of chemical compounds, as conventional machine learning approaches have been more widely explored in this field. Although ANN has achieved remarkable breakthroughs in areas such as computer vision and natural language processing, comparable advances have not yet been fully realized for tabular molecular data [6]. In this work, artificial neural network, a subfield of artificial intelligence (AI), was employed to develop predictive models for classifying pharmaceutical compounds according to their toxicity. To attempt this goal, a tabular dataset with 2,527 molecular observations and 1,545 descriptors have been used [5,7]. An RF-based feature selection method was employed to remove noisy and less relevant features while retaining the most informative descriptors for model development. The respiratory toxicity prediction model was then assessed using both internal and external validation schemes based on widely used classification metrics, including accuracy, precision recall, F1-score, MCC and AUC-ROC.

2 MATERIAL AND METHODS

2.1 Tabular dataset preparation

In this study, a tabular dataset was constructed from molecular data collected through a review of the respiratory toxicity literature, including compounds reported in previous studies (Jaganathan et al., 2022; Wang et al., 2021) [5,7]. A total of 2,527 compounds were collected and classified into two categories: compounds reported to exert adverse effects on the human respiratory system were labeled as toxic, whereas those without reported respiratory toxicity were labeled as non-toxic. Molecular descriptors were subsequently computed using the Mordred and PaDEL libraries implemented in Python 3.9, resulting in a total of 1,545 topological, electronic and physicochemical descriptors for each compound [8-9]. The final tabular dataset was structured with individual compounds as rows and molecular descriptors as columns, along with an additional binary response column encoding respiratory toxicity, where 0 represented non-toxic compounds and 1 represented toxic compounds.

2.2 Feature selection method

At this stage, a Random Forest (RF)-based feature selection approach was employed to identify the most informative molecular descriptors. By recursively partitioning the feature space, RF assesses the contribution of individual features to node impurity reduction and overall predictive performance. Feature importance was therefore used to rank the descriptors and retain the most relevant subset for subsequent model development. The RF model was implemented

using scikit-learn with 1,000 trees (*n_estimators* = 1,000), while all other hyperparameters were kept at their default values [10].

2.3 Artificial neural network implementation

Inspired by the functioning of the human brain, Artificial Neural Networks (ANNs) are computational models designed to mimic, in a simplified manner, the way biological neurons process and transmit information. ANNs are non-parametric models and were employed in this study because of their ability to capture complex and nonlinear relationships within molecular data, thereby providing strong predictive performance. ANNs constitute one of the fundamental architectures underlying deep learning approaches. A typical ANN consists of three main components: an input layer, one or more hidden layers, and an output layer. Input data are introduced into the network through the input layer and propagated through the hidden layers, where successive mathematical transformations are performed, before reaching the output layer to generate the final prediction [11]. Although artificial neural networks (ANNs) have achieved remarkable performance across a wide range of domains, particularly in image processing, computer vision, and natural language processing, their application to tabular data remains challenging. In many real-world settings, data are naturally represented in tabular form, often originating from relational databases and characterized by heterogeneous features and complex relationships that are not always effectively captured by ANN models. Unlike images or text, tabular data consists of structured features that may exhibit heterogeneous distributions and complex relationships, making it particularly difficult to model them effectively using neural network architectures [6]. In this work, an open-source deep learning framework namely PyTorch is used as tool for working with multidimensional arrays called tensors, performing automatic differentiation, defining neural-network architectures, optimizing model parameters, and accelerating computations using GPUs [12].

2.4 Internal and External Model validations

We employed an Artificial Neural Network (ANN) architecture to perform binary classification of pharmaceutical compounds into toxic and non-toxic classes based on their respiratory toxicity profiles. The validation of the developed model was performed using two complementary approaches: internal and external validation. Internal validation was conducted to assess the robustness and stability of the model, whereas external validation was used to evaluate its generalization ability and predictive performance on previously unseen data. A ten-fold cross-validation (10-fold CV) scheme was employed for internal validation. At each fold, a confusion matrix was generated by comparing the predicted and actual toxicity classes. Model performance was evaluated using widely adopted classification metrics, including accuracy (Q), precision(PR), recall (RE), F1-score (F) and Matthews correlation coefficient (MCC), calculated from the resulting true-positive (TP), true-negative (TN), false-positive (FP), and false-negative (FN) counts [13]. In addition, the area under the receiver operating characteristic curve (AUC-ROC) was calculated to evaluate the model's ability to discriminate between toxic and non-toxic compounds across different classification thresholds [14]. Finally, external validation was performed using an independent test set that was excluded from both model training and internal cross-validation. The predictive performance and generalization ability of the finalized ANN model were then assessed on these unseen compounds using the same classification metrics employed for internal validation.

$$Q = \frac{TP + TN}{TP + TN + FP + FN} \quad \text{Eq. 1}$$

$$PR = \frac{TP}{TP + FP} \quad \text{Eq. 2}$$

$$RE = \frac{TP}{TP + FN} \quad \text{Eq. 3}$$

$$F = \frac{2 \times TP}{2 \times TP + FP + FN} \quad \text{Eq.4}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad \text{Eq. 5}$$

2.5 Model explanation

At this stage, SHapley Additive exPlanations (SHAP), a widely used feature-based interpretability method, was employed to explain the model predictions and enhance the transparency and interpretability of the ANN model [15-16]. SHapley Additive exPlanations (SHAP) is an interpretability framework based on concepts from cooperative game theory that provides insights into the predictions of machine learning and deep learning models, which are often considered "*black-box*" approaches. SHAP quantifies the contribution of each input feature to a model prediction, thereby enabling both local interpretation of individual predictions and global assessment of feature importance [16]. This approach can be applied to both supervised regression and classification models. In drug discovery and ADMET profiling, model interpretability is particularly important for understanding how molecular descriptors influence predictions, identifying the most influential descriptors, and determining which molecular features contribute most

significantly to the predicted endpoint. In the present study, SHAP analysis was therefore used to identify and quantify the contribution of molecular descriptors to the ANN predictions of respiratory toxicity.

3 RESULTS AND DISCUSSION

3.1 Random forest for feature selection

A meta-estimator-based feature selection approach was employed as a wrapper method prior to ANN modeling. In this procedure, a Random Forest (RF) model was used as the underlying estimator to evaluate feature importance and identify the most informative molecular descriptors. The selected subset of descriptors was subsequently used as input features for the development of the ANN model. Table 1 summarizes the hyperparameter settings used in the RF-based feature selection procedure to identify the optimal subset of molecular descriptors. The RF model was implemented using the *RandomForestClassifier* available in the scikit-learn Python package [10]. At this point, a feature was considered important if its importance score was greater than the mean importance score across all descriptors. Based on this criterion, 200 features were retained from the initial set of 1,545 molecular descriptors and subsequently used as input variables for ANN model development.

Table 1: Random Forest hyperparameters used for feature selection.

Parameters	Values
N_estimators	1,000
Criterion	gini
Max_depth	none
Min_samples	2.0
Max_feature	sqrt
Min_impurity_decrease	0.0
Bootstrap	True

3.2 Model development

Following feature selection, an Artificial Neural Network (ANN) was developed for the binary classification of pharmaceutical compounds according to their respiratory toxicity using PyTorch framework [12]. Prior to final model development, a predefined hyperparameter grid combined with a cross-validation scheme was employed to optimize the ANN architecture and training parameters. The optimal hyperparameter combination was selected based on cross-validation performance to enhance the predictive ability and generalization of the model. Table 2 summarizes the optimal hyperparameters retained for the final ANN model used to classify pharmaceutical compounds as toxic or non-toxic.

Table 2: ANN hyperparameters used in this study

Parameters	Meaning	Values
Input features	Number of molecular descriptors	200
Hiden Layers	Number of hyden layers	2
Neurons per Hiden layers	Number of neurons in each layer	256, 512
Hiden Activation	Activation function	ReLU
Output Unit	Number of output neurons	1
Output Activation	Classification activation	Sigmoid
Optimizer	Optimization algorithm	Adam
Learning Rate	Optimizer step size	0.001
Batch size	Samples processed per update	256
Epochs	Maximum training epochs	200

Dropout rate	Regularization strength	0.35
Loss Function	Training objective	Binary cross-entropy
Early Stoping	Training termination criterion	yes

3.3 Internal evaluation

At this stage, the tabular dataset used for ANN modeling consisted of 2,527 compounds with 200 descriptors. Using a *train-test split* algorithm, the data were divided into training and test sets, with the *test_size* parameter set to 0.3, indicating that 70% of the data were used for training and 30% for testing. Accordingly, 1,768 molecular observations were used for training, of which 762 compounds were classified as non-toxic and 1,006 compounds as toxic. The ANN model was subjected to a ten-fold cross-validation scheme. After model training, the statistical parameters were evaluated, and the corresponding results are summarized in Table 3. The internal validation statistical parameters demonstrated that the ANN model was robust and showed good predictive performance on tabular data for predicting the toxicity of pharmaceutical compounds.

Table 3: Statistical results of ANN model based on ten-fold cross-validation procedure

Statistical parameters (%)	Values± SD
Q	84.05 ± 2.21
Precision	86.37 ± 2.02
Recall	85.50 ± 2.73
F1-score	85.91 ± 1.97
MCC	67.59 ± 4.49
AUC	91.36 ± 1.55

Accuracy provides an overall assessment of the proportion of pharmaceutical compounds correctly classified by the ANN model, whereas precision evaluates the proportion of compounds predicted as toxic that are actually toxic. Accordingly, an accuracy of 84.05% indicates that approximately 84% of the molecules in the training set were correctly classified by the ANN model. In addition, a precision of approximately 86.37% indicates that ~86% of the compounds predicted as toxic were actually toxic. These results demonstrate good classification performance of the ANN model on the training data. Recall is particularly important in toxicity assessment because a high value indicates a reduced risk of classifying potentially toxic compounds as non-toxic. Indeed, our model demonstrated a high recall value of 85.50%, indicating its good ability to correctly identify compounds belonging to the class of interest. The F1-score provides a balanced measure of precision and recall, while the Matthews correlation coefficient (MCC) assesses the overall agreement between predicted and observed classes by considering all components of the confusion matrix. In addition, the area under the ROC curve (AUC-ROC) evaluates the model's ability to discriminate between toxic and non-toxic compounds across different classification thresholds. Thus, the combined analysis of these statistical parameters provides a more comprehensive assessment of the robustness and predictive performance of the ANN model than accuracy alone. The results of these statistical parameters reported in Table 3 indicate excellent predictive performance of the ANN model on tabular data.

3.4 External evaluation

Regarding the test set, the ANN model achieved an accuracy of 83.40%, indicating that more than 83% of the compounds were correctly classified. The precision of 85.58% indicates that 85.58% of the compounds predicted as toxic were actually toxic, whereas the recall of 85.19% shows that the model correctly identified 85.19% of the truly toxic compounds. The F1-score, which represents the harmonic mean of precision and recall, reached 85.38%, demonstrating a good balance between these two metrics. Furthermore, the MCC value of 0.66 indicates a substantial agreement between the predicted and observed classes while accounting for all components of the confusion matrix. Finally, the AUC value of 89.53% demonstrates a strong ability of the ANN model to discriminate between toxic and non-toxic compounds across different classification thresholds (Fig. 1). Overall, the consistently high performance obtained across these complementary metrics on the independent test set supports the good predictive performance and robustness of the developed ANN model for the classification of pharmaceutical compounds according to their respiratory toxicity.

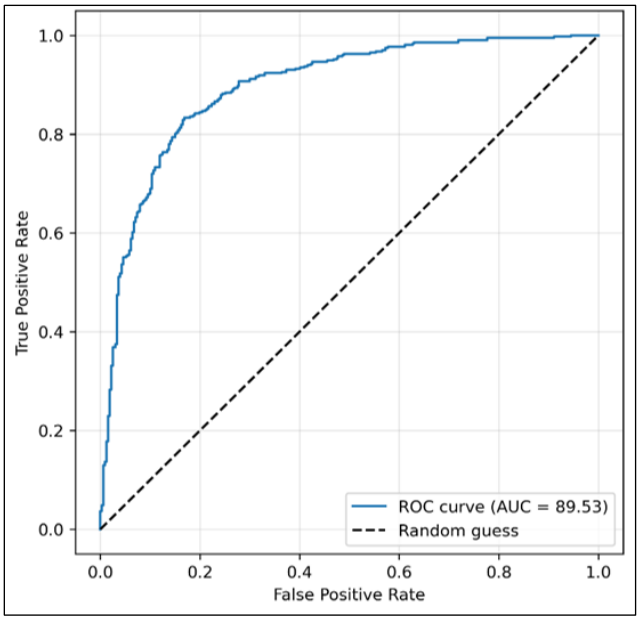
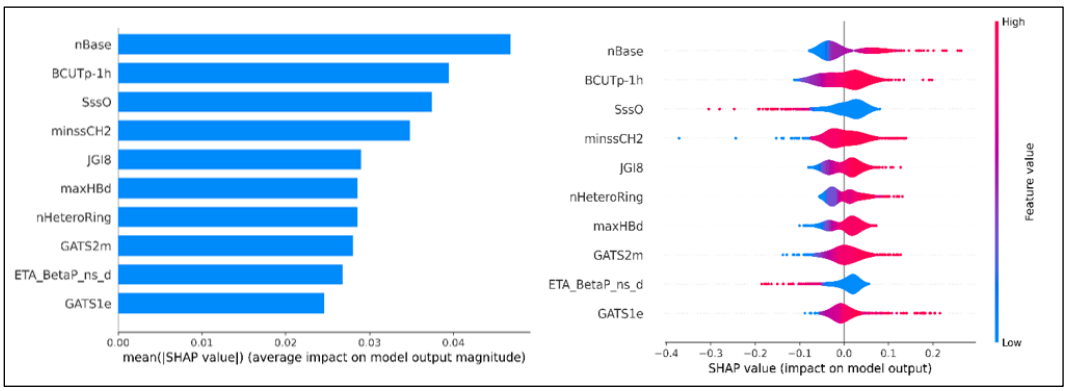


Figure 1: Roc curve and AUC parameters obtained on test set

3.5 ANN model Explanation

SHAP analysis reveals that ANN predictions of pharmaceutical compound-induced respiratory toxicity are driven by a combination of diverse molecular determinants. In particular, nBase descriptor reflects the number of basic groups and therefore provides information on molecular basicity and potential ionization behavior. Interestingly, the importance of this descriptor is consistent with a previous explainable machine-learning study of pulmonary toxicity, in which nBase emerged as the most influential molecular descriptor and higher values were associated with an increased likelihood of classification as a pulmonary toxicant [5]. As shown in Figure 2, the molecular descriptors identified by SHAP as the most influential to model predictions encompass several complementary descriptor classes including constitutional and functional-group descriptors (nBase and nHeteroRing), electrotopological-state descriptors (SssO, minssCH2, and maxHBd), BCUT descriptors (BCUTp-1h), topological charge indices (JGI8), 2D autocorrelation descriptors (GATS2m and GATS1e), and extended topochemical atom descriptors (ETA_BetaP_ns_d) [10-11]. Collectively, these descriptors encode information related to molecular basicity, heterocyclic structure, local electronic environments, hydrogen-bonding characteristics, atomic polarizability, charge distribution, atomic mass, electronegativity, and molecular topology. Their contribution to the ANN predictions therefore suggests that pharmaceutical compound-induced respiratory toxicity is associated with a combination of structural, electronic, and physicochemical molecular characteristics rather than with a single molecular property.



(a) (b)

Figure 2: (a) Mean SHAP values (b) Impact on model output

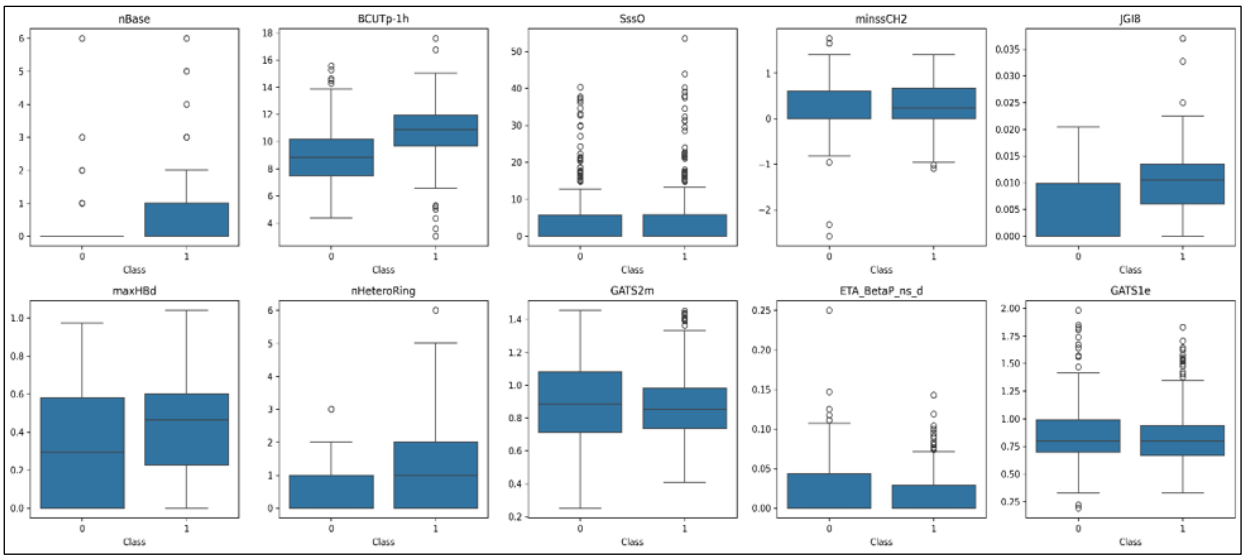


Figure 3: Distribution of descriptors selected through SHAP analysis

To further characterize the molecular features highlighted by the SHAP analysis, the distributions of the most influential descriptors were examined according to respiratory toxicity class (0: non-toxic compounds; 1: toxic compounds). As no statistical tests were performed to assess differences in means or distributions between the two classes, this analysis was considered descriptive and was primarily intended to provide complementary information for interpreting the SHAP results. The selected descriptors showed varying distributional patterns across toxic and non-toxic compounds, with differences in central tendency, dispersion, and degree of overlap observed for some descriptors (fig. 3). These patterns provide additional context for understanding the contribution of the descriptors identified as important by SHAP, without implying statistically significant differences between the two toxicity classes. Descriptors showing visually distinct distributional patterns may help support their relevance to the ANN predictions, whereas substantial overlap between classes may suggest that their predictive contribution arises from nonlinear relationships or interactions with other molecular features rather than from simple univariate class separation. Overall, the combined examination of descriptor distributions and SHAP-based explanations provides a more comprehensive interpretation of the structural and physicochemical characteristics underlying the ANN predictions of pharmaceutical compound-induced respiratory toxicity.

3.6 Study limitations

Despite its promising predictive performance, this study has several limitations. The model was developed from 2,527 pharmaceutical compounds using mainly 2D descriptors, which may not fully capture three-dimensional, metabolic, and biological factors involved in respiratory toxicity. Moreover, although RF-based feature selection was applied, no specific procedure was used to systematically address descriptor multicollinearity, which may influence SHAP feature attribution. While ten-fold cross-validation and test set validation procedures supported model robustness, validation using an independent external dataset is still required. In addition, SHAP identifies associations with model predictions rather than causal relationships, and the distributions of important descriptors were examined descriptively without statistical testing. Finally, experimental validation remains necessary to confirm the predicted toxicity and associated molecular determinants.

4 CONCLUSION

In this study, an artificial neural network (ANN) model was developed to predict pharmaceutical compound-induced respiratory toxicity from tabular data of molecular descriptors. Model performance was assessed using ten-fold cross-validation and external validation based on multiple statistical metrics, including accuracy, precision, recall, F1-score, MCC, and AUC-ROC. The ANN showed robust predictive performance, achieving accuracies of 84.05% and 83.40% during internal and external validation, respectively, with an AUC-ROC of 89.53% on the test set. These results demonstrate the model's ability to discriminate between toxic and non-toxic pharmaceutical compounds while maintaining satisfactory generalization to unseen data. To improve model transparency and gain insight into the molecular properties associated with toxicity predictions, SHAP analysis was applied to the ANN model. Collectively, these molecular descriptors capture complementary structural, electronic and physicochemical characteristics, including molecular topology, polarizability, hydrogen-bonding capacity, local electronic environments, and the spatial distribution of atomic physicochemical properties. The SHAP results therefore suggest that the ANN predictions of

respiratory toxicity arise from the combined contribution of multiple structural and physicochemical characteristics rather than from a single molecular property.

STATEMENTS ON COMPLIANCE WITH ETHICAL STANDARDS

Acknowledgments

The authors gratefully acknowledge Nangui Abrogoua University and the members of the Laboratory of Fundamental and Applied Physics (LFAP) for their institutional, scientific, and collegial support.

Disclosure of conflict of interest

The authors declare no conflicts of interest regarding this manuscript.

REFERENCES

[1] van der Merwe, D. (2025). Chapter 11 - Respiratory toxicity. Dans R. C. Gupta (Éd.), Veterinary Toxicology (Fourth Edition) (éd. Fourth Edition, pp. 185-196). Academic Press. doi:https://doi.org/10.1016/B978-0-443-29007-7.00042-X

[2] Koirala M, Yan L, Mohamed Z, DiPaola M. AI-Integrated QSAR Modeling for Enhanced Drug Discovery: From Classical Approaches to Deep Learning and Structural Insight. Int J Mol Sci. 2025 Sep 25;26(19):9384. doi: 10.3390/ijms26199384.

[3] N'guessan A, Akonan L, MeledjeD, N'guessan H, Allangba GP, Mousse L, et al. An Efficient Random Forest Model for Predicting Respiratory Toxicity of Organic Chemicals. J. Pharm. Res. Int. [Internet]. 2025 Nov. 15 ;37(12):1-10. https://doi.org/10.9734/jpri/2025/v37i127774

[4] Jaganathan, K., Geethika, P.R., Ramakrishnan, S. et al. Explainable AI and tree-based ensemble models: a comparative study in predicting chemical pulmonary toxicity. Eur. Phys. J. Spec. Top. 234, 2659–2671 (2025). https://doi.org/10.1140/epjs/s11734-024-01291-3

[5] Jaganathan, K.; Tayara, H.; Chong, K.T. An Explainable Supervised Machine Learning Model for Predicting Respiratory Toxicity of Chemicals Using Optimal Molecular Descriptors. Pharmaceutics 2022, 14, 832. https://doi.org/10.3390/pharmaceutics14040832

[6] Hegselmann, S., Buendia, A., Lang, H., Agrawal, M., Jiang, X., & Sontag, D. (2023, April). Tabllm: Few-shot classification of tabular data with large language models. In International conference on artificial intelligence and statistics, PMLR, 5549-5581.

[7] Wang, Z., Zhao, P., Zhang, X., Xu, X., Li, W., Liu,G., & Tang, Y. (2021). In silico prediction of chemical respiratory toxicity via machine learning. Computational Toxicology, 18,100155.

[8] Yap CW. PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. J Comput Chem. 2011 May;32(7):1466-74. doi: 10.1002/jcc.21707. Epub 2010 Dec 17. PMID: 21425294.

[9] Moriwaki, H., Tian, YS., Kawashita, N. et al. Mordred: a molecular descriptor calculator. J Cheminform 10, 4 (2018). https://doi.org/10.1186/s13321-018-0258-y

[10] Pedregosa, F. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825

[11] Qamar, Roheen and Zardari, Baqar Ali (2023) Artificial Neural Networks: An Overview, Mesopotamian Journal of Computer Science, 3(1) DOI: https://doi.org/10.58496/MJCSC/2023/015

[12] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury et al. (2019). Pytorch: An imperative style, high-performance deep learning library, Advances in neural information processing systems, 32.

[13] Tharwat, A. (2021). Classification assessment methods. Applied computing and informatics, 17(1), 168-192.

[14] Fawcett, T. (2006). An introduction to ROC analysis. Pattern recognition letters, 27(8), 861-874.

[15] Ponce-Bobadilla AV, Schmitt V, Maier CS, Mensing S, Stodtmann S. Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development. Clin Transl Sci. 2024 Nov;17(11):e70056. doi: 10.1111/cts.70056. PMID: 39463176; PMCID: PMC11513550.

[16] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in neural information processing systems, 30.