Innov Biosyst Bioeng, 2025, vol. 9, no. 2, 42–52
doi: 10.20535/ibb.2025.9.2.316239

Original article

IN SILICO THE AMES MUTAGENICITY PREDICTIVE MODEL OF ENVIRONMENT

S.V. Kislyak[1*], O.M. Duhan[1], R.V. Yesypenko[1], D.B. Starosyla[2], O.I. Yalovenko[1]

[1]Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, Ukraine
[2]George Washington School of Medicine and Health Sciences, Washington DC, USA

Corresponding author: kisluak@ukr.net

Received 16 December 2024; Accepted 2 April 2025

Background. The classical *in vitro* and *in vivo* methods developed and widely used in the past decades to assess the genetic effects of environmental factors are complex in view of their implementation, are expensive, long-lasting, have the problem of reproducibility of the results of experiment in different laboratories and may face ethical problems of using warm-blooded animals in experiments.
Objective. Development, optimisation and testing of effective *in silico* models for assessment of Ames mutagenicity of environmental factors.
Methods. The genetic assessment of the impact of environmental factors was carried out in accordance with a set of chemical compounds for which information on potential mutagenic activity was obtained experimentally, using the *in vitro* Ames Salmonella/microsome test. Four machine learning models were developed to solve the problem of binary classification to form two classes of xenobiotics (mutagen/non-mutagen). The total sample is represented by a set of 8,083 xenobiotics.
Results. We developed four machine learning models with 85% accuracy, matching the reproducibility of Ames test data across laboratories. In addition, we have proposed a binary classifier that subject to dimensionality reduction of the input data, taking into account the qualitative composition of molecular descriptors, allows us to improve the accuracy of *in silico* prediction of genotoxicity of chemicals.
Conclusions. The necessity of updating and expanding the list of effective and more productive methods and approaches for assessing the genotoxic effects of environmental factors is substantiated, which allows avoiding the use of warm-blooded animals in the experiment, saving time and reducing the number of false-negative and false-positive results. The possibility of increase the accuracy of predictive machine learning models for assessing the genotoxic potential of environmental factors in conditions of dimensionality reduction of the data set is presented.

Keywords: mutation; genotoxicity; QSAR model; molecular descriptors; machine learning models.

Introduction

Over the past few decades, due to scientific and technological progress, there has been an exponential increase in the number of chemical compounds released into the environment that can affect the human genetic apparatus. In August 2024, the number of registered xenobiotics, information on which is stored on the servers of the American Chemical Society, was more than 280 million substances. As of the beginning of 2020, information was available on more than 100,000 chemicals produced by industry that can adversely affect the environment and human health, and, in particular, genetic health [1].

One of the main features of chemical hazards is their ability to interact with hereditary material, which can initiate the development of various genetic and oncological diseases [2–5]. Despite the fact that the issue of genetic and carcinogenic safety is receiving a lot of attention, the public and the scientific community concern about the fact that information on the specific biological effects of a large number of compounds is, on the one hand, insufficiently studied, and on the other hand, the results of studies on such xenobiotics have been either contradictory or obtained with significant limitations on the final genetic effects. Another significant drawback is the fact that a large number of chemicals (especially medicines, various food additives, cosmetics, household chemicals) are registered without conducting researches on genetic assessment of impact on the environment and the human genetic apparatus [6–8]. Significant limitations in the assessment of genetic effects of potential genotoxic compounds are associated with the necessity to take into account the main provisions of the "3R" concept *in vivo tests*, which is guided by the principles aimed at reducing, improving and replacing animal models [9, 10]. The standardised *in vitro* [11–14] and *in vivo* [15–20] genotoxicity assessment methods approved by the Organisation

for Economic Co-operation and Development and accepted by the scientific community in conditions of the extremely rapid expansion of the chemical space do not allow us to obtain information on the mutagenic potential of a large number of substances – potential mutagens – released into the environment [21, 22]. In addition, the drawbacks of the standard battery testing systems [23, 24] for determining genotoxicity also include the complexity of testing, the rather high cost of research, and the ethical problems of using warm-blooded animals, which contradict the provisions of the "3R" concept. The problem of reproducibility of the experiment is related to the diversity of batches of culture media, as well as the subjective factor, which is a major obstacle to obtaine objective results in assessing the genotoxic potential of environmental factors.

Among the various *in vitro* and *in vivo* experimental methods for genetic assessment of the impact of environmental factors, *in vitro* the Ames test deserves special attention. The Ames test is used by almost all laboratories around the world to detect and assess the genetic activity of environmental factors. In addition, for more than 50 years of using this method, a sufficiently large amount of experimental material has been obtained as to the number of chemical compounds tested, which can be used to develop more efficient and modern *in silico* models for assessing the Ames mutagenicity of environmental factors. In accordance with the recommendations of the Organisation for Economic Co-operation and Development, five strains of *Salmonella typhimurium* are used to assess mutagenic potential (TA1535, TA1537, TA98, TA100, TA102) [11]. The increased interest of the scientific community in this method of genotoxicity testing is also due to the "Guidelines on Genotoxicity Testing and Data Interpretation for Pharmaceuticals for Human Use" ICH S2(R1) adopted in 2012 by the International Council on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, which presented two theoretically justified schemes for the use of *in vitro* and *in vivo* batteries of standard testing systems for genotoxicity determination, each of which included the Ames test [25].

The development of information technologies and artificial intelligence systems, as well as advances in computer science, computational molecular biology and chemoinformatics, have become a reliable foundation for further activation of the scientific community to resolve the urgent issues of modern computational toxicology. The result of these processes became a significant paradigm shift in genotoxicity testing, which is associated with the development, testing and implementation of modern *in silico* QSAR models for assessing the genotoxic effects of environmental factors. The abbreviation QSAR (Quantitative Structure-Activity Relationship) is used in the scientific literature to denote models of quantitative structure-activity relationship. Within the task solution to obtain a genetic assessment of the impact of environmental factors, the abbreviation QSAR combines methods that allow us to predict the activity (mutagenicity) of a xenobiotic according to its structure, which is given by a set of molecular descriptors. The increased interest of scientists in the development of effective *in silico* models can be traced in scientific papers [26–30]. Despite the sufficient attention of researchers to the creation of QSAR models for determining genotoxic potential, the number of scientific papers in this area of research has not decreased over the past 5 years, but has even increased. This trend is determined by the high predictive potential of *in silico* models. The Ames/QSAR predictive models that use the *in vitro* Ames test results for various xenobiotics as input data deserve special attention. An interesting fact is that for the Ames/QSAR models described in the scientific literature, the problem of improving the quality of classification for chemical compounds that may exhibit potential genotoxic properties needs to be resolved. In this situation, the prospects for improving the predictive capability of *in silico* models can be realised by covering a larger number of chemical compounds, using various balanced sets of molecular descriptors, 2-D digital structure impress, applying long-term predictive QSAR models and implementing approaches to optimise them.

Taking into account the sufficiently large potential of machine learning models and the peculiarities of the input data used to obtain a genetic assessment of the impact of environmental factors, the most quality models in view of predicting of mutagenicity, the Ames/QSAR models can be obtained by reducing the set of molecular descriptors and selecting those predictors that have a significant impact on the predicted variable.

Materials and Methods

In creating of *in silico* machine learning models, we used a dataset [31] obtained by combining three publicly available datasets: Kazius-Bursi [32], Hansen [33] and EFSA [34]. According to the dataset containing 8,083 xenobiotics, any chemical

compound was considered mutagenic if at least one positive result was obtained in the *in vitro* the Ames test on *S. typhimurium* strains TA97, TA98, TA100, TA102, TA1535, TA1537 and TA1538. For the Ames/QSAR predictive models, a set of 1,442 descriptors for each potential environmental pollutant was used, obtained in accordance with the linear SMILES notation, using the PaDEL software [31].

Data pre-processing for Ames/QSAR machine learning models involves a standard procedure of their normalisation taking into account the standard range of values from [0.1] and z-normalisation, which allows us to remove anomalous values. At the stage of data preparation, the column "Canonical SMILES", which contains textual information about the structure of the molecule that is not used for modelling, was removed. Columns with identical values were also removed as this information would not be informative for machine learning models. In order to avoid multicollinearity and reduce the data dimensionality, correlated features were also removed when two or more features have a high correlation (more than 0.95) according to the Pearson correlation coefficient. In *in silico* machine learning models, the used data was divided into training and test sets in the ratio of 75/25, respectively.

Four machine learning models have been proposed to solve the binary classification problem: logistic regression (LR-Scikit), logistic regression using stochastic gradient descent (LR-SGD), random forest method, and neural network.

For the logistic regression (Scikit-learn library), most of the model setup parameters were used by default. The problem of unbalanced representation of two classes of xenobiotics (mutagen/non-mutagen), which may affect the quality of classification of the less represented class, was solved by setting class_weight="balanced". At the same time, weighting factors were calculated for each class to ensure an equal impact at the model training stage.

Logistic regression (LR-SGD) uses the stochastic gradient descent (SGD) method to optimise the parameters of the Ames/QSAR model. The ease of SGD implementation is the basis for expanding the horizons of its application in machine and deep learning models [35]. The maximum performance of SGD is achieved by updating the model parameters on each individual data sample or a small group of samples (batches). In our proposed Ames/QSAR LR-SGD machine learning model, 64 samples from the training dataset were used to update the model parameters. The logistic regression model was implemented through an architecture based on the *Sequential* class of the *TensorFlow* library, which allows us to create models consisting of a sequence of layers. To solve the binary classification problem, we chose a simple architecture that included only one (*Dense*) layer with one output. After passing through the *Dense* layer, the weights and offsets are calculated based on the input data. The model receives input data in the form of a matrix, where each row represents an observation and each column represents a feature. The *sigmoid* was chosen as the activation function, which converts a linear combination of features into a probability that a certain chemical compound belongs to one of two classes – mutagen or non-mutagen. The process of effective training model involves optimization the loss function, which minimises the difference between the predicted probability value and the actual result of the mutagenic potential assessment of a particular xenobiotica. To solve the problem of binary classification in the context of genotoxic potential assessment, it was proposed to use the loss function *binary_crossentropy*.

The *RandomForestClassifier* class of the *Scikit-learn* Pyton library was used to implement the Ames/QSAR random forest model. The random forest model was represented by 200 trees, (*n_estimators=200*), with a maximum number of leaves equal to 600 (*max_leaf_nodes=600*).

The structure of the neural network consists of an input layer, an output layer, and 4 hidden layers containing 128, 256, 128, and 64 neurons, respectively. The *ReLU* function was selected as the activation function for the hidden layers, which has the advantages of relative simplicity of implementation and allows us to effectively solve the problem of the vanishing gradient. At the output layer, the activation function is *Sigmoid*, which makes it possible to obtain the probability of xenobiotics belonging to one of two classes (mutagen/non-mutagen). To solve the standard neural network problem connected with the *L1* and *L2* methods and *Dropout* regularisation were used. In the created neural network, *L1* regularisation is used on the second hidden layer, and *L2* regularisation is used on the third hidden layer. There is a *Dropout* between all layers of the network, which randomly switches off 30–60% of neurons during the training. To minimise the loss function, the most efficient optimiser was chosen based on the adaptive estimation of the Adam moment (adaptive moment estimation) [36]. The neural network training was carried out for 100 epochs with a batch size of 64.

The performance of the developed *in silico* the Ames/QSAR predictive models was evaluated using the metrics of accuracy, recall, specificity and F_1-score, which were calculated taking into account the confusion matrix in accordance with the relations

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN},$$

$$recall = \frac{TP}{TP + FN},$$

$$specificity = \frac{TN}{TN + FP},$$

$$F_1\text{-}score = 2 \cdot \frac{precision \cdot recall}{precision + recall},$$

where TP, TN, FP, FN correspond to the number of true positive, true negative, false positive and false negative classification results, respectively.

Similar approaches to the selection of Ames/QSAR predictive models for solving the problem of assessing the genotoxicity of environmental factors can be traced in scientific papers [22, 27, 37, 38]. At the stage of selection the best machine learning model, we also used a popular and quite effective metric based on calculating the area under the receiver operating characteristic curve (ROC) [39]. The performance of the four predictive models was evaluated according to the number of *True Positives*, *True Negatives*, *False Positives* and *False Negatives* results in the classification. It should be noted that the accuracy metric is considered in view of assessing two parameters: overall *accuracy*, which corresponds to the part of xenobiotics that were correctly distributed between the two classes, and *precision* metric, which corresponds to the part of chemical compounds that were correctly classified by the Ames/QSAR predictive models as chemical compounds with pronounced genotoxic properties. The *recall* metric allows us to determine the proportion of chemical compounds that are mutagens, taking into account the total number of true positive and false negative results. The criterion for evaluating the performance of Ames/QSAR *specificity* models is similar to *recall*, but is calculated to determine the part of xenobiotics that are negative for genotoxic potential, taking into account the total number of true negative and false positive classification results. F_1-score, as a criterion for evaluating the performance of developed machine learning models, is a harmonic average of two metrics – *precision* and *recall*.

Results

Table 1 presents the classification results obtained on the test sample using LR-Scikit, LR-SGD, random forest method, and neural network.

According to obtained classification reports, taking into account *accuracy*, *recall*, *specificity* metrics and F_1-*score*, among the four predictive Ames/QSARs, the best, taking into account all metrics, is the Random Forest method with AUC = 0.92. The neural network demonstrated less performance with an accuracy of 0.83 and a sensitivity of 0.87. The area under the ROC curve for the neural network is 0.9, which indicates a fairly good prediction result of the classification model. Despite the fact that the values of the classification reports for LR-Scikit and LR-SGD are almost the same, the analysis of error matrices allowed us to give preference to the latter method. On the test sampling, the LR-SGD model, in comparison with LR-Scikit, allowed us to identify a larger number of true positive xenobiotics that show genotoxicity properties.

The optimisation of machine learning models by reducing the dimensionality of the input data was implemented using the generated ranked list of molecular descriptors, which was obtained according to the coefficients of two regression models (LR-SGD and LR-Scikit) and two Random Forest methods: mean decrease impurity and permutation feature importance. Table 2 presents a list of molecular descriptors that were identified according to the four approaches, occurred several times in the models, and had a significant enough impact on the predicted variable. These model parameters are important in view of assessment of classification

Table 1: Classification results for the test sample

Metrics	LR-Scikit	LR-SGD	Random Forest	Neural network
accuracy	0.79	0.79	0.86	0.83
recall	0.81	0.83	0.87	0.82
specificity	0.76	0.76	0.84	0.83
F_1-*score*	0.80	0.80	0.86	0.83

Notes. LR-Scikit – logistic regression, LR-SGD – logistic regression using stochastic gradient descent.

Table 2: List of molecular descriptors that have a significant impact on the predicted variable

	Random Forest (mean decrease impurity)	Random Forest (permutation feature importance)	LR-Scikit	LR-SGD
1	*MATS1e*	**SHBint2**	*AATS2s*	**GATS1p**
2	**R_TpiPCTPC**	*MATS1e*	**GATS1p**	*nAtomP*
3	*nFRing*	**BCUTp-1h**	**R_TpiPCTPC**	*SpMax1_Bhm*
4	*nAtomP*	*ATSC2e*	**BCUTp-1h**	**R_TpiPCTPC**
5	*MLFER_E*	*SpMin4_Bhm*	*nHBd*	*AATSC2i*
6	*SpMin1_Bhm*	*SpMin1_Bhm*	*AATSC0m*	*nFRing*
7	**GATS1p**	*GATS1m*	*AATSC2i*	*MATS2c*
8	*GATS1m*	*AATSC0m*	*MATS2c*	*AATS2s*
9	*ATSC2e*	–	*SpMax1_Bhm*	*nHBd*
10	–	–	**SHBint2**	**BCUTp-1h**
11	–	–	*ATS7s*	*ATS7s*
12	–	–	*MLFER_E*	**SHBint2**
13	–	–	*SpMin4_Bhm*	–

Notes. LR-Scikit – logistic regression, LR-SGD – logistic regression using stochastic gradient descent.

results. Given the large number of molecular descriptors used as input data for the developed classifiers, it was decided to form a limited list of the most important descriptors (Table 2), taking into account the thirtieth molecular descriptor of the ranked list. Molecular descriptors that were presented in one of the models only once were not recorded in Table 2. Within this study, we paid attention to the descriptors that may have a significant impact on the manifestations of xenobiotic genotoxicity. We also used the ranked list of descriptors to solve the problem of optimising machine learning models, which consisted of determining a fixed base set of features that are selected taking into account the list of descriptors recorded in descending order of weighting. At the same time, both recurring descriptors and those that were represented in the models once were taken into account.

Table 2 presents in bold the mnemonics of molecular descriptors that are repeated in the models 3-4 times. The molecular descriptors that are repeated in only two models are in italics (see Table 2). The order of recording the molecular descriptors for each method corresponds to the values of the weighting coefficients (modulo), which are decreased with increasing row number in Table 2.

It is scientifically important to effectively solution of the problem of classification for potential genotoxic compounds using predictive models, taking into account a limited number of molecular descriptors. This approach is the basis for improving the models, allowing obtaining a better generalizability and resilience of the models to retraining [40]. To solve the problem of the Ames/QSAR models optimisation, we proposed using as input data for each of the four classifiers a set of 50, 100, 150, 200, 300, and 400 molecular descriptors selected from each model. The features were selected according to a ranked list of molecular descriptors obtained by taking into account the coefficients of two regression models (LR-SGD and LR-Scikit) and two Random Forest methods: mean decrease impurity and permutation feature importance. The removal of duplicate descriptors affected the final value of the number of descriptors that were further used in the modelling. The number of features used for testing the Ames/QSAR models without duplicates was 128 (was 200), 276 (was 400), 371 (was 600), 454 (was 800), 589 (was 1,200) and 655 (was 1,600). At the same time, the duplicate molecular descriptors were saved in a separate file for further testing *in silico* models with a limited data set, represented by features with the highest weighting coefficients and having a significant impact on the predicted variable.

Table 3 presents information on the accuracy assessment of the classification results for the four machine learning models obtained on the test sampling for a fixed number of molecular descriptors without duplicates.

According to the results of the machine learning model testing, there is no significant correlation between the accuracy of the models and the qualitative and quantitative composition of molecular descriptors. Reducing the number of molecular descriptors usually led to a decrease in the accuracy of *in silico* models. It is quite interesting from a scientific point of view that the use of a data set with 276 molecular descriptors without duplicates led to an improvement in the predictive ability of logistic regressions and a neural network. At the same time, the accuracy of the random forest model did not change (the Figure).

Table 4 presents the classification results obtained on the test sampling using LR-Scikit, LR-SGD, random forest method, and neural network, taking into account a limited data set of 276 molecular descriptors.

According to the obtained classification reports, we can note an increase in the accuracy of logistic regressions, which increased to 80%, and the neural network – to 84 %.

To solve the issue related to the selection of this set of input data, in which the efficiency of the developed predictive models will be the best, we proposed to conduct additional testing of machine learning models taking into account a set of molecular duplicate descriptors that have a significant impact on the predicted variable. To evaluate the performance of the developed classifiers, 78 molecular duplicate descriptors were used, which were removed at the stage of formation of input data sets different in quantity (see Table 3). The classification result for the test sampling with a limited number of duplicate descriptors is presented in Table 5.

Table 3: Dependence of the accuracy on Ames/QSAR models on the qualitative and quantitative composition of molecular descriptors

Number of descriptors (without repetitions)	Accuracy of LR Scikit-learn	Accuracy of LR-SGD	Accuracy of RF	Accuracy of NN
128	78%	77%	84%	82%
276	80%	80%	86%	84%
371	78%	78%	84%	80%
454	78%	78%	83%	82%
589	79%	79%	84%	82%
655	80%	79%	83%	84%
1442 (initial value)	79%	79%	86%	83%

Notes. LR-Scikit – logistic regression, LR-SGD – logistic regression using stochastic gradient descent, RF – random forest, NN – neural network.

Table 4: Classification results for the test sampling (276 molecular descriptors)

Metrics	LR-Scikit	LR-SGD	Random Forest	Neural network
accuracy	0.80	0.80	0.86	0.84
recall	0.80	0.81	0.84	0.85
specificity	0.79	0.80	0.88	0.83
F_1*-score*	0.80	0.80	0.86	0.84

Notes. LR-Scikit – logistic regression, LR-SGD – logistic regression using stochastic gradient descent.

Table 5: Classification result for the test sampling with a limited number of duplicate molecular descriptors

Metrics	LR-Scikit	LR-SGD	Random Forest	Neural network
accuracy	0.74	0.74	0.77	0.78
recall	0.73	0.72	0.80	0.84
specificity	0.75	0.76	0.85	0.72
F_1*-score*	0.74	0.74	0.82	0.80

Notes. LR-Scikit – logistic regression, LR-SGD – logistic regression using stochastic gradient descent.

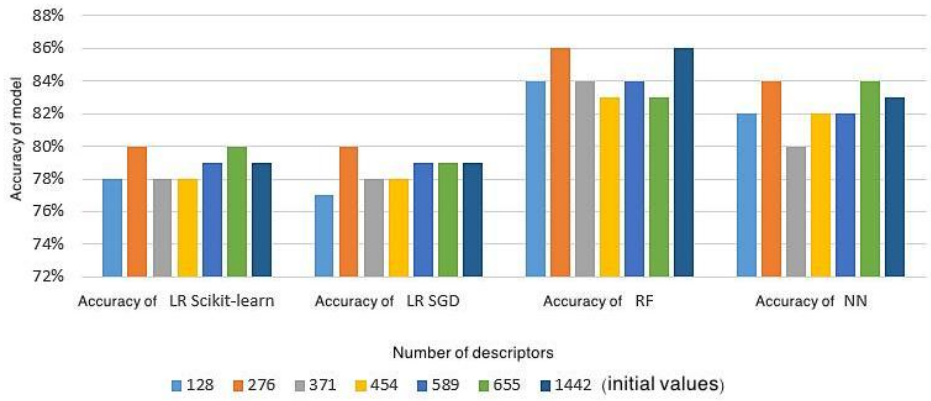

Figure: Histogram of Ames/QSAR models accuracy dependence on the quantitative composition of molecular descriptors

According to the obtained classification reports, taking into account the *accuracy*, *recall*, *specificity,* and F_1-*score* metrics, we can see a decrease in the efficiency of the Ames/QSAR models developed compared to the classification results with the full set of input data (for 1,442 molecular descriptors). The result of testing machine learning models with a limited set of molecular descriptors has some negative implications in view of assessment of their performance. On the other hand, reducing the number of molecular descriptors by 95 % of the original number resulted in a decrease in accuracy for the four predictive models of only from 3 to 6 %. This result can act as a stimulus for the search for cause-and-effect relationships between mutagenicity and physicochemical, spatial, electronic characteristics of a particular xenobiotic, which are given by a set of molecular descriptors-duplicates that have a significant impact on the predicted variable.

Discussion

The study of the impact of environmental factors on the human genetic apparatus is one of the priority areas of modern genetic toxicology. The classical scheme for detecting and assessing the genotoxic potential of chemical compounds involves the use of a standard battery of *in vitro* and *in vivo* test systems, which has significant defects in view of time spent, cost of experimental studies and ethical problems [23, 24, 29, 41]. Trends in the development of modern computational toxicology, which should comply with the basic principles of the "3R" concept, require a reduction in the number of studies with warm-blooded animals. In conditions of the increasing number of chemicals in the environment that can show genotoxic properties, scientists pay special attention to *in silico* models that can act as alternative approaches for genetic assessment of environmental factors. The use of *in silico* QSAR models is a promising forward-looking approach to solving the classification problem for a set of chemicals with uncertain genotoxic potential and will optimise the complex process of effective identification and accounting of chemical compounds that may impact on the human genetic apparatus. The trend towards the active use of machine learning algorithms and the introduction of effective *in silico* methods for genotoxicity assessment can be traced in scientific papers [22, 27, 37, 38, 42–44]. This intensification of the scientific community is explained by the rather large unrealised potential of machine learning models and the active implementation of new approaches that can be crucial in terms of improving the efficiency of the developed models.

Despite the advantages and prospects of using *in silico* machine learning models in toxicology, they have certain defects. The main obstacle to creating effective predictive Ames/QSAR models is the availability of a sufficiently large number of chemical compounds that are protected by security documents and cannot usually be used in modelling. In addition, the predictive ability of the developed models directly depends on the quality of the experimental data obtained on *S. typhimurium* strains (TA1535, TA1537, TA98, TA100, TA102).

The accuracy of modern Ames/QSAR described in the scientific literature is currently in the range of 80–85%, which corresponds to the reproducibility of the Ames test in different laboratories [45, 46]. At the same time, the accuracy of the four Ames/QSAR models developed by us, taking into account a limited set of 276 molecular descriptors, ranged from 80 to 86 %.

The assessment of the genotoxic potential of xenobiotics that may be presented in the environment is usually based on a sufficiently large number of molecular descriptors [37, 47]. The optimisation of the machine learning models developed in this paper was realised by reducing the amount of input data used for modelling. To solve this problem, we used the coefficients of two regression models (LR-SGD and LR-Scikit) and two Random Forest methods: mean decrease impurity and permutation feature importance, which allowed us to select the most influential features with the subsequent formation of a ranked list of molecular descriptors that should be taken into account when conducting *in silico* modelling. This approach has scientific value and may allow further researchstudy to resolve the issue of finding cause and effect relationships between genotoxicity and features defined by a set of molecular descriptors. In addition, the formation of a list of basic sets of molecular essential descriptors (see Table 2) may allow us to obtain a higher quality Ames/QSAR predictive model. The relevant methodology of using a limited set of molecular descriptors can also be applied to different classes of genotoxic compounds, which may further simplify the classification task solving chemical compounds with similar physical and chemical properties. A detailed analysis of molecular descriptors (see Table 2) in relation to the physicochemical, spatial, electronic features of a particular xenobiotic is quite interesting from a scientific point of view. First of all, we focused on a set

of molecular descriptors that are repeated 3-4 times (see Table 2). A rather interesting result was obtained after analysing scientific sources on the use of duplicate molecular descriptors in similar studies with the implementation of predictive QSAR models. The molecular descriptor GATS1p, which encodes information on molecular polarizability, was used as the main one in predicting chemical toxicity for 1,163 agents. At the same time, the toxicity of xenobiotics was assessed using QSAR models on the test microorganism *Tetrahymena pyriformis*. Quite interesting is the fact that out of 753 molecular descriptors, only 4 were selected in the modelling, including GATS1p [48]. The R_TpiPCTPC descriptor was also selected as the main descriptor in the Ames/QSAR model used to assess mutagenicity, carcinogenicity and genotoxicity in mammals [49]. SHBint2 is the main descriptor that was selected in the implementation of two machine learning models for determining the Ames mutagenicity. At the same time, three predictive models were implemented in the study [50]. The molecular descriptor BCUTp-1h, which is one of the criteria for assessing a xenobiotic molecule as to the level of intermolecular interactions, belongs to the main descriptor of one of the QSAR models proposed by the authors [51], which was developed to assess the toxicity of pesticides dissolved in water. The obtained results add confidence that the formation of a list of important features, given by a set of molecular descriptors, can be carried out in accordance with the methodology presented in this paper.

Conclusions

Among the four implemented Ames/QSAR predictive models (taking into account the full set of 1,442 molecular descriptors and 8,083 chemical compounds), the best one is the Random Forest method with an AUC = 0.92 and an accuracy rate of 86%. The neural network demonstrated lower efficiency with 83% accuracy and 82% sensitivity. Comparison of classification reports for the two regressions allows us to give preference to the LR-SGD, which allows us to effectively identify a larger number of true positive chemical compounds in view of mutagenicity. The developed Ames/QSAR predictive models are balanced in terms of identifying both positive (mutagenic) and negative (non-mutagenic) cases, as defined by the *recall* and *specificity* metrics.

The optimisation of Ames/QSAR models, which was implemented by reducing the amount of input data, was realised by using a list of molecular descriptors ranked according to the weighting coefficients. This procedure, on the one hand, allowed us to develop a methodology for formation a list of the main descriptors that have a significant impact on the predicted variable and should be taken into account in modelling. On the other hand, the generated list of descriptors was used to identify a limited set of descriptors that would increase the predictive ability of the developed models. The conducted study showed that the use of a limited set of 276 molecular descriptors led to increase in accuracy for logistic regressions, which increased to 80%, and for the neural network – to 84%. At the same time, the accuracy of the Random Forest model remained unchanged and was 86%. This result indicates that when developing Ames/QSAR models, it is necessary to take into account not only the basic, molecular descriptors that can be obtained with the maximum values of the weighting coefficients, according to the methodology developed by us, but also a set of features that may have a minor impact on the predicted variable. In this situation, increasing the accuracy of the Ames/QSAR models used to assess the genotoxic effects of environmental factors can be achieved by combining of less influential descriptors with other. The solution to the problem associated with finding the best set of features for which the Ames/QSAR models developed by the authors will be most effective may be in search for cause and effect relationships between mutagenicity and physicochemical, spatial, electronic characteristics of a particular xenobiotic, which are defined by a set of molecular descriptors.

Modern trends in the development of artificial intelligence and approaches that are the basis for deep learning give hope for solving all the pressing issues in the context of creating effective *in silico* models for predicting Ames mutagenicity.

Interests disclosure

The authors declare no conflicts of interests.

References

[1] Honma M. An assessment of mutagenicity of chemical substances by (quantitative) structure-activity relationship. Genes Environ. 2020 Jul 2;42:23. DOI: 10.1186/s41021-020-00163-1

[2] Burgess JT, Rose M, Boucher D, Plowman J, Molloy C, Fisher M, et al. The therapeutic potential of DNA damage repair pathways and genomic stability in lung cancer. Front Oncol. 2020 Jul 28;10:1256. DOI: 10.3389/fonc.2020.01256

[3] Milanese C, Cerri S, Ulusoy A, Gornati SV, Plat A, Gabriels S, et al. Activation of the DNA damage response in vivo in synucleinopathy models of Parkinson's disease. Cell Death Dis. 2018 Jul 26;9(8):818. DOI: 10.1038/s41419-018-0848-7

[4] Waller R, Murphy M, Garwood CJ, Jennings L, Heath PR, Chambers A, et al. Metallothionein-I/II expression associates with the astrocyte DNA damage response and not Alzheimer-type pathology in the aging brain. Glia. 2018 Nov;66(11):2316-23. DOI: 10.1002/glia.23465

[5] Wang H, Li S, Oaks J, Ren J, Li L, Wu X. The concerted roles of FANCM and Rad52 in the protection of common fragile sites. Nat Commun. 2018 Jul 18;9(1):2791. DOI: 10.1038/s41467-018-05066-y

[6] Yalovenko OI, Ostanina NV, Dugan OM. Tinted balms in the system of staged testing of their potential mutagenic effect on genetically modified microorganisms. Cytol Genet. 2005;39(6):60-5.

[7] Yalovenko OI. Assessment of the biological action of persistent hair dyes using alternative test systems. Environ Health. 2006;1:58-63.

[8] Yalovenko OI, Raetska OV, Maistrenko ZY., Dugan OM. Assessment of mutagenic activity of individual ingredients and finished products (perfume, cosmetic, detergent) in alternative biological test systems. Environ Health. 2005;2:63-5.

[9] Chen L, Li N, Liu Y, Faquet B, Alépée N, Ding C, et al. A new 3D model for genotoxicity assessment: EpiSkin™ Micronucleus Assay. Mutagenesis. 2021 Apr 28;36(1):51-61. DOI: 10.1093/mutage/geaa003

[10] Chung YH, Gulumian M, Pleus RC, Yu IJ. Animal welfare considerations when conducting OECD test guideline inhalation and toxicokinetic studies for nanomaterials. Animals (Basel). 2022 Nov 26;12(23):3305. DOI: 10.3390/ani12233305

[11] Organisation for Economic Co-operation and Development (OECD). Test No. 471: Bacterial Reverse Mutation Test. In: OECD Guidelines for the Testing of Chemicals, Section 4. Paris: OECD Publishing; 2016. DOI: 10.1787/9789264071247-en

[12] Organisation for Economic Co-operation and Development (OECD). Test No. 473: In Vitro Mammalian Chromosomal Aberration Test. In: OECD Guidelines for the Testing of Chemicals, Section 4. Paris: OECD Publishing; 2016. DOI: 10.1787/9789264264649-en

[13] Organisation for Economic Co-operation and Development (OECD). Test No. 487: In Vitro Mammalian Cell Micronucleus Test. In: OECD Guidelines for the Testing of Chemicals, Section 4. Paris: OECD Publishing; 2023. DOI: 10.1787/9789264264861-en

[14] Organisation for Economic Co-operation and Development (OECD). Test No. 489: In Vivo Mammalian Alkaline Comet Assay. In: OECD Guidelines for the Testing of Chemicals, Section 4. Paris: OECD Publishing; 2016. DOI: 10.1787/9789264264885-en

[15] Organisation for Economic Co-operation and Development (OECD). Test No. 474: Mammalian Erythrocyte Micronucleus Test. In: OECD Guidelines for the Testing of Chemicals, Section 4. Paris: OECD Publishing; 2016. DOI: 10.1787/9789264264762-en

[16] Organisation for Economic Co-operation and Development (OECD). Test No. 475: Mammalian Bone Marrow Chromosomal Aberration Test. In: OECD Guidelines for the Testing of Chemicals, Section 4. Paris: OECD Publishing; 2016. DOI: 10.1787/9789264264786-en

[17] Organisation for Economic Co-operation and Development (OECD). Test No. 476: In Vitro Mammalian Cell Gene Mutation Tests using the Hprt and xprt genes. In: OECD Guidelines for the Testing of Chemicals, Section 4. Paris: OECD Publishing; 2016. DOI: 10.1787/9789264264809-en

[18] Organisation for Economic Co-operation and Development (OECD). Test No. 478: Rodent Dominant Lethal Test. In: OECD Guidelines for the Testing of Chemicals, Section 4. Paris: OECD Publishing; 2016. DOI: 10.1787/9789264264823-en

[19] Organisation for Economic Co-operation and Development (OECD). Test No. 485: Genetic toxicology, Mouse Heritable Translocation Assay. In: OECD Guidelines for the Testing of Chemicals, Section 4. Paris: OECD Publishing; 2016. DOI: 10.1787/9789264071506-en

[20] Organisation for Economic Co-operation and Development (OECD). Test No. 486: Unscheduled DNA Synthesis (UDS) Test with Mammalian Liver Cells in vivo. In: OECD Guidelines for the Testing of Chemicals, Section 4. Paris: OECD Publishing; 1997. DOI: 10.1787/9789264071520-en

[21] Lou C, Yang H, Deng H, Huang M, Li W, Liu G, et al. Chemical rules for optimization of chemical mutagenicity via matched molecular pairs analysis and machine learning methods. J Cheminform. 2023;15(1):35. DOI: 10.1186/s13321-023-00707-x

[22] Shinada NK, Koyama N, Ikemori M, Nishioka T, Hitaoka S, Hakura A, et al. Optimizing machine-learning models for mutagenicity prediction through better feature selection. Mutagenesis. 2022;37(3-4):191-202. DOI: 10.1093/mutage/geac010

[23] Luan Y, Honma, M. Genotoxicity testing and recent advances. Genome Instab Dis. 2022;3:1-21. DOI: 10.1007/s42764-021-00058-7

[24] Mohamed S, Sabita U, Rajendra S, Raman D. Genotoxicity: mechanisms, testing guidelines and methods. Glob J Pharm Sci. 2017;1(5):555575. DOI: 10.19080/GJPPS.2017.01.555575

[25] Food and Drug Administration, HHS. International Conference on Harmonisation; guidance on S2(R1) Genotoxicity Testing and Data Interpretation for Pharmaceuticals intended for Human Use; availability. Notice. Fed Regist. 2012 Jun 7;77(110):33748-9.

[26] Lui R, Guan D, Matthews S. Mechanistic task groupings enhance multitask deep learning of strain-specific ames mutagenicity. Chem Res Toxicol. 2023 Aug 21;36(8):1248-54. DOI: 10.1021/acs.chemrestox.2c00385

[27] Li T, Liu Z, Thakkar S, Roberts R, Tong W. DeepAmes: A deep learning-powered Ames test predictive model with potential for regulatory application. Regul Toxicol Pharmacol. 2023 Oct;144:105486. DOI: 10.1016/j.yrtph.2023.105486

[28] Kalian AD, Benfenati E, Osborne OJ, Gott D, Potter C, Dorne JCM, et al. Exploring dimensionality reduction techniques for deep learning driven QSAR models of mutagenicity. Toxics. 2023 Jun 30;11(7):572. DOI: 10.3390/toxics11070572

[29] Guo W, Liu J, Dong F, Song M, Li Z, Khan MKH, et al. Review of machine learning and deep learning models for toxicity prediction. Exp Biol Med (Maywood). 2023 Nov;248(21):1952-73. DOI: 10.1177/15353702231209421

[30] Kislyak S, Duhan O, Yalovenko O. Systems for genetic assessment of the impact of environmental factors. Innov Biosyst Bioeng. 2024 Mar 25;8(2):3-27. DOI: 10.20535/ibb.2024.8.2.288127

[31] Helma C, Schöning V, Drewe J, Boss P. A Comparison of nine machine learning mutagenicity models and their application for predicting pyrrolizidine alkaloids. Front Pharmacol. 2021 Jul 22;12:708050. DOI: 10.3389/fphar.2021.708050

[32] Kazius J, McGuire R, Bursi R. Derivation and validation of toxicophores for mutagenicity prediction. J Med Chem. 2005 Jan 13;48(1):312-20. DOI: 10.1021/jm040835a

[33] Hansen K, Mika S, Schroeter T, Sutter A, ter Laak A, Steger-Hartmann T,et al. Benchmark data set for in silico prediction of Ames mutagenicity. J Chem Inf Model. 2009 Sep;49(9):2077-81. DOI: 10.1021/ci900161g

[34] EFSA Panel on Contaminants in the Food Chain (CONTAM). Scientific Opinion on Pyrrolizidine alkaloids in food and feed. EFSA Journal 2011;9(11):2406. DOI: 10.2903/j.efsa.2011.2406

[35] Du J. The frontier of SGD and its variants in machine learning. J Phys Conf Ser. 2019;1229(1):012046. DOI: 10.1088/1742-6596/1229/1/012046

[36] Kingma DP, Ba JA. A method for stochastic optimization. In: 3rd International Conference for Learning Representations; 2015 May 7-9; San Diego. DOI: 10.48550/arXiv.1412.6980

[37] Fan D, Yang H, Li F, Sun L, Di P, Li W, et al. In silico prediction of chemical genotoxicity using machine learning methods and structural alerts. Toxicol Res (Camb). 2017 Dec 15;7(2):211-20. DOI: 10.1039/c7tx00259a

[38] Nguyen-Vo TH, Nguyen L, Do N, Le PH, Nguyen TN, Nguyen BP, et al. Predicting drug-induced liver injury using convolutional neural network and molecular fingerprint-embedded features. ACS Omega. 2020;5(39):25432-9. DOI: 10.1021/acsomega.0c03866

[39] Nahm FS. Receiver operating characteristic curve: overview and practical use for clinicians. Korean J Anesthesiol. 2022 Feb;75(1):25-36. DOI: 10.4097/kja.21209

[40] Guyon I, Elisseeff A, Kaelbling LP, editors. An introduction of variable and feature selection. J Machine Learn Res. 2003;3(708)1157-82. DOI: 10.1162/153244303322753616

[41] Barylyak IR, Dugan OM. Ecological and genetic studies in Ukraine. Cytol Genet. 2002;5:3-10.

[42] Sofuni T. Evolution of genotoxicity test methods in Japan. Genes Environ. 2017;39:15. DOI: 10.1186/s41021-016-0063-7

[43] Furukawa A, Ono S, Yamada K, Torimoto N, Asayama M, Muto S. A local QSAR model based on the stability of nitrenium ions to support the ICH M7 expert review on the mutagenicity of primary aromatic amines. Genes Environ. 2022 Mar 21;44(1):10. DOI: 10.1186/s41021-022-00238-1

[44] Rane R, Satpute B, Kumar D, Suryawanshi M, Prabhune AG, Gawade B, et al. Mutagenic and genotoxic *in silico* QSAR prediction of dimer impurity of gliflozins; canagliflozin, dapaglifozin, and emphagliflozin and *in vitro* evaluation by Ames and micronucleus test. Drug Chem Toxicol. 2025 Mar;48(2):416-25. DOI: 10.1080/01480545.2024.2378768

[45] Honma M, Kitazawa A, Cayley A, Williams RV, Barber C, Hanser T, et al. Improvement of quantitative structure-activity relationship (QSAR) tools for predicting Ames mutagenicity: outcomes of the Ames/QSAR International Challenge Project. Mutagenesis. 2019 Mar 6;34(1):3-16. DOI: 10.1093/mutage/gey031

[46] Yang X, Zhang Z, Li Q, Cai Y. Quantitative structure-activity relationship models for genotoxicity prediction based on combination evaluation strategies for toxicological alternative experiments. Sci Rep. 2021;11(1):8030. DOI: 10.1038/s41598-021-87035-y

[47] Benigni R, Bassan A, Pavan M. In silico models for genotoxicity and drug regulation. Expert Opin Drug Metab Toxicol. 2020 Aug;16(8):651-62. DOI: 10.1080/17425255.2020.1785428

[48] Yu X. Prediction of chemical toxicity to Tetrahymena pyriformis with four-descriptor models. Ecotoxicol Environ Saf. 2020 Mar 1;190:110146. DOI: 10.1016/j.ecoenv.2019.110146

[49] Guan D, Fan K, Spence I, Matthews S. QSAR ligand dataset for modelling mutagenicity, genotoxicity, and rodent carcinogenicity. Data Brief. 2018 Feb 2;17:876-84. DOI: 10.1016/j.dib.2018.01.077

[50] Votano JR, Parham M, Hall LH, Kier LB, Oloff S, Tropsha A, et al. Three new consensus QSAR models for the prediction of Ames genotoxicity. Mutagenesis. 2004 Sep;19(5):365-77. DOI: 10.1093/mutage/geh043

[51] Galimberti F, Moretto A, Papa E. Application of chemometric methods and QSAR models to support pesticide risk assessment starting from ecotoxicological datasets. Water Res. 2020 May 1;174:115583. DOI: 10.1016/j.watres.2020.115583

С.В. Кисляк[1], О.М. Дуган[1], Р.В. Єсипенко[1], Д.Б. Старосила[2], О.І. Яловенко[1]

[1]КПІ ім. Ігоря Сікорського, Київ, Україна
[2]Факультет медицини і охорони здоров'я Університету Джорджа Вашингтона, Вашингтогн, США

IN SILICO МОДЕЛІ ПРОГНОЗУВАННЯ МУТАГЕННОСТІ ЕЙМСА ФАКТОРІВ НАВКОЛИШНЬОГО СЕРЕДОВИЩА

Проблематика. Розроблені та широко використовувані в минулі десятиріччя класичні *in vitro* та *in vivo* методи оцінки генетичних ефектів факторів навколишнього середовища є складними з точки зору їх проведення, є дороговартісними, тривалими в часі, мають проблему відтворюваності результатів експерименту в різних лабораторіях і можуть стикатися з етичними проблемами використання в експериментах теплокровних тварин.
Мета. Розробка, оптимізація й апробація ефективних *in silico* моделей оцінки мутагенності Еймса впливу факторів навколишнього середовища
Методика реалізації. Генетична оцінка впливу факторів навколишнього середовища була проведена відповідно до набору хімічних сполук, для яких експериментально, за допомогою *in vitro* тесту Еймса Salmonella/microsome, була отримана інформація про потенційну мутагенну активність. Для розв'язання задачі бінарної класифікації з метою формування двох класів ксенобіотиків (мутаген/не мутаген) було розроблено чотири моделі машинного навчання. Загальну вибірку, що представлена набором із 8083 ксенобіотиків, було розділено на тренувальну та валідаційну у співвідношенні 75 до 25% відповідно.
Результати. Точність розроблених моделей машинного навчання була в межах 85%, що відповідає відтворюваності експериментальних даних, отриманих у кількісному, напівкількісному та якісному тестах Еймса в різних лабораторіях. Запропоновано бінарний класифікатор, що за умов зменшення розмірності вхідних даних дає змогу підвищити точність результатів *in silico* прогнозування мутагенності Еймса.
Висновки. Обґрунтовано необхідність оновлення та розширення переліку ефективних і більш продуктивних методів і підходів для оцінки генотоксичних ефектів факторів навколишнього середовища, що дає змогу уникнути застосування в експерименті теплокровних тварин, заощадити час та зменшити кількість хибнонегативних і хибнопозитивних результатів. Показано можливість збільшення точності прогностичних моделей машинного навчання для оцінки генотоксичного потенціалу впливу факторів навколишнього середовища за умов зменшення розмірності набору даних.

Ключові слова: мутація; генотоксичність; QSAR-модель; молекулярні дескриптори; моделі машинного навчання.