

Received March 20, 2022, accepted April 9, 2022, date of publication April 13, 2022, date of current version April 20, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3167050

Efficient Compound Activity Prediction Model Based on Bayesian Regularization Method for Candidate Drugs

ZHANGMIAOGE LIU^{ID}, NING WANG^{ID}, (Member, IEEE), CHEN ZHANG^{ID}, AND ZHOUXIAO LIU

Shanghai Key Laboratory of Modern Optical System, Engineering Research Center of Optical Instrument and System, Ministry of Education, University of Shanghai for Science and Technology, Shanghai 200093, China

Corresponding author: Ning Wang (nwang@usst.edu.cn)

This work was supported in part by the Shanghai Natural Science Foundation under Grant 21ZR1446300, in part by the Key Laboratory of Wide Band-gap Semiconductor Materials, Ministry of Education, Xidian University, 710071, China (Program No.kdxkf2020-02), in part by the Natural Science Basic Research Program of Shaanxi (Program No.2021JM-384), in part by the NSFC under Grant 61804096, and in part by the National Key Research and Development Program of China under Grants 2018YFA0701800.

ABSTRACT In breast cancer diagnoses, antagonist compounds are used to resist α Subtype of Estrogen Receptor (ER α) bioactivity. However, those compounds are difficult to be obtained in the process of drug screening. In this paper, an efficient bioactivity prediction model is proposed with consideration of Absorption, Distribution, Metabolism, Excretion and Toxicity (ADMET) properties as an auxiliary prediction to achieve the minimization of the damage caused to human body whereas maximizing the efficacy of estrogen activity inhibitors. In the proposed prediction model, Pearson correlation analysis and Bayesian regularization algorithm under neural network are specifically used to train and analyze the data of bioactivity. To calculate the loss of neural network, cross-entropy and supervise learning are applied. The results show that the accuracy of our proposed compound prediction model can reach up to 92.2%, 94.3%, and 90.2% for Caco2, CYP3A4, and hERG, respectively. Moreover, the Area Under the Curve (AUC) value of our proposed model is 0.9752, which is 13-14% higher than that of SVM method. In addition, the Matthews Correlation Coefficient (MCC) can reach up to 0.93 ± 0.03 , 37-40% higher than that of SVM method. Due to correlation analysis and valid data screening, our model can shorten the prediction time by 15%. Due to its advantages in accuracy and speed of prediction, this research can contribute to the upgrading of diagnosis and treatment of breast cancer.

INDEX TERMS Deep neural networks, drug discovery, machine learning, virtual screening.

I. INTRODUCTION

Breast cancer is one of the most common and highly lethal cancers in the world, whose development is closely related to Estrogen Receptor [1]–[3]. Research has found that Estrogen Receptors alpha (ER α) is expressed in no more than 10% of normal breast epithelial cells but about 50-80% of breast tumor cells. Based on the feature, antihormone therapy is commonly used by breast cancer patients. It controls estrogen level by regulating estrogen receptor activity [4], [5]. Therefore, compound which can resist ER α bioactivity is considered as an important target for treating breast cancer. To save time and cost in drug discovery, compound activity prediction models are usually used to screen potential active compounds. Before building a model, the expected

compounds need to possess some basic features, such as good bioactivity, pharmacokinetic properties and nontoxicity in human body which are collectively known as ADMET. No matter how active a compound is, if its ADMET properties do not meet the requirements, for example, if it was difficult to be absorbed by the human body, or the metabolic rate in the body was too fast, or it had biotoxicity, it would be difficult to be made into medicine, thus the ADMET properties need to be accurately evaluated.

Previously, scientists examined molecules of a substance in the laboratory to determine its effects on human bodies or animals. However, it is hard for them to check all the compounds in the database because it is highly time-consuming and expensive. To solve this problem, virtual screening was created, which is a method that can match known compounds with unknown ones. One of the traditional techniques is similarity search [6]. It is a ligand-based technology, which can

The associate editor coordinating the review of this manuscript and approving it for publication was Marco Giannelli^{ID}.

discover the similarity between molecules or the substructure of distance molecules, but it was only applied within a limited scope. In recent years, machine learning technology has been widely used in the field of chemical informatics with a lot of methods being developed, such as Naive Bayes (NB), Support Vector Machine (SVM), Single hidden Layer Feedforward Network (SLFN), Logistic Regression (LR), and Deep Neural Networks (DNN) [7]–[9]. When using SVM method, the model trained with large-scale samples often has the defect of slow speed, and is sensitive to missing data, parameters and kernel functions [10]. Hence, the traditional SVM is not suitable for multi-classification. Meanwhile, apart from the low accuracy, too much iterative calculation and tuning need to be consumed for SLFN, leading to a long training time [11]. In this regard, Kudisthalert *et al.* discovered Extreme Learning Machine (ELM) prediction model, using single-layer feedforward neural network structure. It selected the weight from the training randomly and processed the data through a generalized inverse function, which can produce a function with small training error. Without performing iterative process to train a model, the training time for ELM can be remarkably reduced compared with SLFN. Yet this model is not widely used due to the complexity of function calculation of approximate hidden layer nodes. These models are usually measured by two main parameters, namely Area Under the Curve (AUC) and Matthews Correlation Coefficient (MCC) [12]–[14].

In this paper, aiming at the high complexity of drug screening, Pearson correlation analysis is used to screen the training data. In dealing with the insufficiency of accuracy of the drug prediction, a quantitative prediction model of compound bioactivity is constructed. A three-layer neural network under Bayesian regularization algorithm is established for effectively training and analyzing the bioactivity data of 1942 compounds. In the meantime, considering the wellbeing of human bodies, a classification prediction for predicting ADMET properties is added to the model, and the loss of neural network model are calculated by cross entropy and supervised learning.

II. METHOD AND PRINCIPLE

In this part, we have introduced the method and principle of our proposed model. Including noise-filtering neural network, the noise filtering of training data, and the Bayesian Regularization Algorithm.

A. NOISE-FILTERING NEURAL NETWORK

To promote the efficient screening of drugs and train a practical drug-active prediction system, we constructed the prediction model shown in Figure 1. In our work, the whole prediction model is simplified and described as two parts, pre-processing and post-processing. The “pre-processing” evaluates the correlation of all variables affecting the prediction results through correlation operation and Bayesian analysis. The dependent variables which have low correlation with the prediction results and constant in all training data are

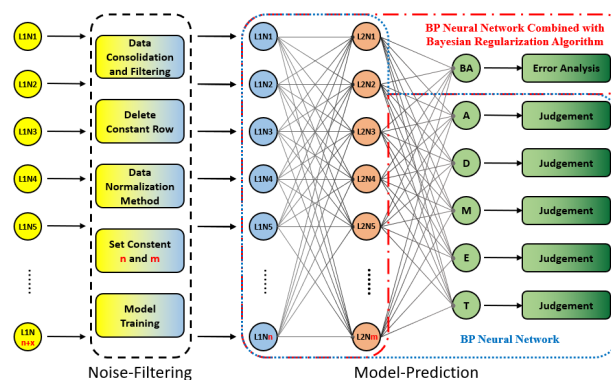


FIGURE 1. The diagram of noise-filtering neural network combined with Bayesian regularization algorithm.

eliminated to filter the noise, reduce the complexity of model training, and improve the accuracy. “Post-processing” refers to that our prediction model based on Bayesian regularization and three-layer neural network is constructed to train the relationship between molecular descriptors and biological activities of known compounds. The number of nodes in the middle layer can first obtain an approximate value through empirical formula, and then obtain an optimal value through adjusting parameters in a certain range. The weights of each layer of neural network nodes are trained by supervised learning. And then the unknown compounds are predicted and compared with the real value to verify the effectiveness and practicability of our proposed model for drug screening.

For Biological Activity (BA) prediction, the BP neural network and Bayesian regularization algorithm are combined to significantly improve the generalization ability, convergence speed and approximation accuracy of BP neural network for the best BA prediction effect. For the simplified ADMET prediction (de facto two classification prediction), the conventional BP neural network is used for classification prediction. In Figure 1, “L1Nn” refers to layer 1 node n, supposing n nodes are in layer 1. Similarly, “L2Nm” represents layer 2 node m, assuming there are m nodes altogether in layer 2. “BA” indicates biological activity, “A, D, M, E, and T” represents five properties to be tested, whose performance can be judged by the output.

$$h = \sqrt{m + n} + a \quad (1)$$

The number of hidden layer nodes has an impact on the performance of neural network. An empirical formula can determine the number of hidden layer nodes, as shown in formula (1), in which h is the number of hidden layer nodes, m is the number of input layer nodes, n is the number of output layer nodes, and a is the adjustment constant between 1 and 10 [15].

After obtaining the number of hidden layer nodes, the ADMET properties begin to predict, as a parameter to measure the comprehensive index of drugs, simplified as 0 and 1 which means inferior and superior. The steps of predicting this parameter are divided into three steps: normalization

processing (formula 2), weight matrix calculation and summation (formula 3) and activation function output (formula 4 and formula 5). The second layer repeats a similar operation of the first layer.

$$(x + \text{offset}) * \text{gain} + y_{\min} = x_{p1} \quad (2)$$

$$S_j = \sum_{i=0}^{m-1} w_{ij} x_i + b_j \quad (3)$$

$$x_j = f(S_j) \quad (4)$$

$$a = \frac{2}{1 + \exp(-2 * n)} - 1 \quad (5)$$

In formula (3), W_{ij} represents the weight between node i and node j , whereas b_j acts for the threshold of node j , and x_j stands for the output value of each node. The output value of each node should be realized according to the output value of all nodes in the upper layer, the weight of the current node and all nodes in the upper layer, the threshold of the current node and the activation function [16], [17].

B. NOISE FILTERING OF TRAINING DATA

Pearson correlation coefficient is often used to calculate the correlation between variables. From a mathematical point of view, it is defined as “the covariance between two vectors is normalized by the product of their standard deviation” [18]–[20]. The covariance between two pairs of vectors is the measurement of their fluctuation trend up and down the mean. That is, it measures whether a pair of vectors tend to be on the same side or opposite to their respective averages [21]–[23]. The covariance is calculated by subtracting the respective mean value from each pair of variables, and then multiplying the two values, as shown in formula (6).

$$\text{Cov}(x, y) = \frac{\sum_i^N (x_i - \bar{x})(y_i - \bar{y})}{N - 1} \quad (6)$$

$$\rho_{xy} = \frac{\text{Cov}(x, y)}{\sigma_x \sigma_y} \quad (7)$$

To obtain more reliable numbers, normalized covariance can be obtained by dividing covariance by the product of the standard deviation of two vectors [24]–[26]. As shown in formula (7), the Greek letter ρ is used to express Pearson correlation coefficient. As shown in equation (8), the covariance of two identical vectors is also equal to their variance. Therefore, the maximum value of the covariance between two vectors is equal to the product of their standard deviation, which occurs when the vectors are completely correlated [27]. The correlation coefficient is limited to $[-1, 1]$.

$$\text{Cov}(x, x) = \text{Var}(x) \quad (8)$$

$$R(i, j) = \frac{C(i, j)}{\sqrt{C(i, i) \cdot C(j, j)}} \quad (9)$$

According to the correlation analysis, the Correlation Coefficient function is mainly used to calculate the correlation of variables by Pearson correlation algorithm, as shown in formula (9). Where, $C(i, j)$ is the covariance matrix,

$R(i, j)$ is the correlation coefficient matrix, and the correlation coefficient is calculated from the covariance.

C. BAYESIAN REGULARIZATION ALGORITHM

When the size of the training sample set is certain, the generalization ability of the network is directly related to the scale of the network. If the size of the neural network is much smaller than that of the training sample set, the chance of “over training” would be very small. However, for solving specific problems, it is often difficult to determine the appropriate network size (the number of hidden layer neurons). The regularization method improves the generalization ability of neural network by modifying the training performance function of neural network. In general, the training performance function of neural network is expressed by formula (10) [28], [29].

$$E_D = \frac{1}{N} \sum_{i=1}^N (e_i)^2 = \frac{1}{N} \sum_{i=1}^N (t_i - a_i)^2 \quad (10)$$

$$E = \zeta_1 \cdot E_D + \zeta_2 \cdot E_W \quad (11)$$

$$E_W = \frac{1}{N} \sum_{j=1}^N W_j^2 \quad (12)$$

In the regularization method, the mean square sum of network weights is added to the typical objective function, and the network performance function is improved to the formula (11) and formula (12) with ζ_1 and ζ_2 being the parameters. If $\zeta_1 \gg \zeta_2$, the training algorithm minimizes the network error. Conversely, if $\zeta_1 \ll \zeta_2$, the training emphasizes the reduction of weight, leading to a relatively great network error, which eventually makes the output of the neural network smoother and prevents the occurrence of minimum points [30].

By using the new performance index function, the network is adjusted to a small weight and the effective weight of the network is minimized under the condition of ensuring that the network training error is as small as possible, that is equivalent to automatically reducing the scale of the network. It is difficult to determine the scale coefficient (ζ_1 and ζ_2) by conventional regularization methods, whereas the Bayesian regularization method adjusts the weight and threshold of the network according to the Levenberg Marquardt optimization theory, that adaptively can adjust the parameters (ζ_1 and ζ_2) of the objective function in the network training and make it optimal [31].

III. RESULTS AND DISCUSSION

In this part, we have shown the results of correlation analysis, bioactivity prediction, and ADMET properties prediction, respectively. Then, we have discussed these results.

A. CORRELATION ANALYSIS

For obtaining objective and accurate prediction results, we have downloaded bioactivity data tables and molecular descriptor information tables of 1974 compounds from

PubChem database. The bioactivity data training table contains three columns. The first column provides the structural formulas of 1974 compounds, represented by SMILES (Simplified Molecular Input Line Entry System); the second column is the bioactivity value of compounds, who have impacts on ER α ; the third column is pIC50, which is usually positively correlated with bioactivity. In the molecular descriptor information table, the first column is the SMILES formula of the compound, followed by 729 columns, and each column represents a molecular descriptor of the compound.

The correlation coefficients between each molecular descriptor variable and bioactivity variable are calculated according to Correlation Coefficient function, with the coefficients displayed from left to right according to the list number, as shown in Figure 2. As shown in the figure, the number of variables is as high as 729, and the amount of data obtained is relatively large. It is impossible to intuitively see the 20 variables with the largest correlation coefficient. The molecular descriptors with the most significant effect on biological activity can be obtained by rows sorting function. Due to correlation analysis and valid data screening, our proposed model can shorten the prediction time by 15%.

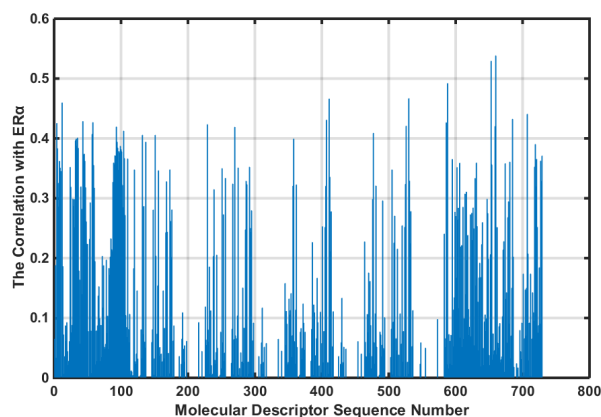


FIGURE 2. Distribution of variable correlation coefficient.

B. PREDICTION OF BIOLOGICAL ACTIVITY

In this part, the network parameters are set as follows: the number of network epochs is set as 100, the expected error goal as 0.00001, and the learning rate LR as 0.1. After being tried, it is safe to say that when the number of hidden layer nodes is 14, the test accuracy of the test set is the highest at 86%, hence the number of hidden layer nodes is set to 14. Then the neural network model trained is tried firstly under Levenberg Marquardt (LM) algorithm, the Relevance (R) value is obtained as 0.77761 and Mean-Square Error (MSE) is 0.72564. It is apparent that it cannot meet the expectation in the MSE. For comparing with other training method, the Bayesian algorithm is used to train the neural network. The given 1974 data sets are divided into two parts, with the first 1800 data sets trained by Bayesian method, and the remaining 174 data sets being tested.

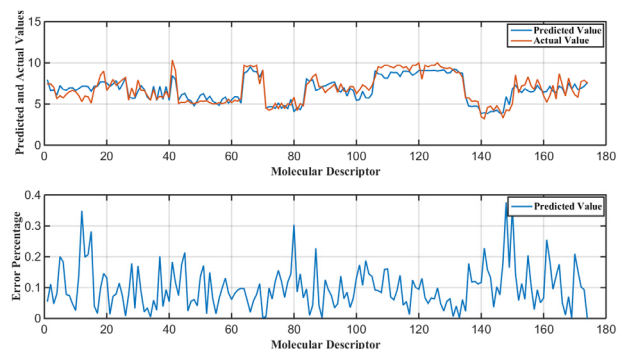


FIGURE 3. Analysis image of predicted value, real value and error.

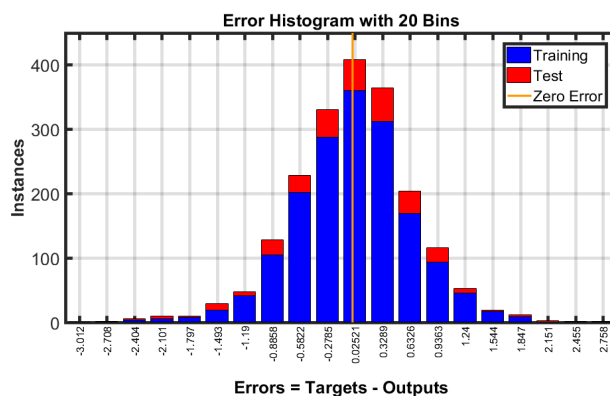


FIGURE 4. Error distribution square image of training set and test set.

As the predicted results shown in Figure 3, to test the error between the real value and the predicted value of 174 molecular descriptor variables in the test set and their error percentage, the red broken line and blue broken line are correspondingly the actual value and predicted value of biological activity of each molecular descriptor. In the figure, the test values and the actual values are showing a similar trend, the error percentage of nearly 93% of the data is controlled within a range of 3%, proving that the established model can make an accurate prediction of the data.

Figure 4 shows the error distribution histogram of training set and test set. The blue column and red column represent the number of training sets and test sets with errors within this range respectively. It is not difficult to see that the training error and the test set error are concentrated in $[-0.2785, 0.3289]$, and only a few data errors have absolute values greater than 1.5.

Figure 5 is the relationship image of the MSE value of the training set data and the number of test sets with the increase of the number of epochs, in which the red curve and the blue curve respectively are the MSE values of the test data set and the training data set. In the first two epochs, the MSE decreased rapidly from 10 to 1, and from the 81st epoch, the test MSE and training MSE of the model basically remained at 0.40184. Compared with the MSE value of LM algorithm, the MSE under Bayesian algorithm has been greatly reduced.

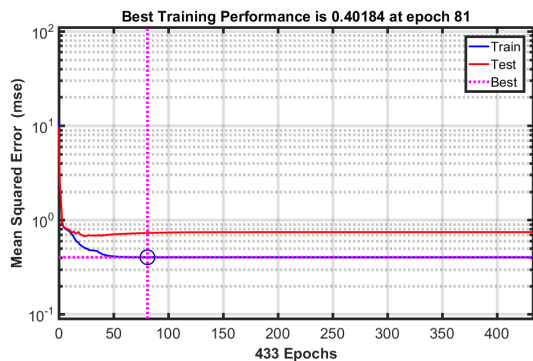


FIGURE 5. Variation of MSE of training set, test set and overall with training times.

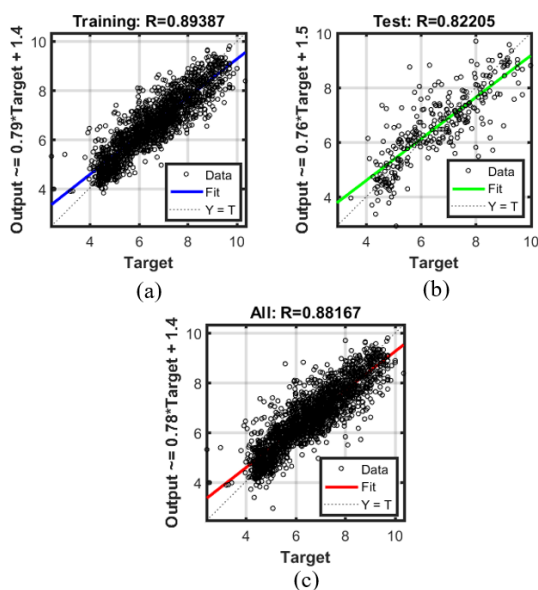


FIGURE 6. (a). Correlation analysis images of training set. (b). Correlation analysis images of test set. (c). Correlation analysis images of all dataset.

Figure 6 shows the correlation distribution images of each sample in the training set and the test set on the proposed model. The training set data correlation R is 0.89387, whereas the test set data correlation R is 0.82205, and the overall correlation R reached 0.88167. As a result, the accuracy of our proposed model trained by Bayesian method on the test data set is higher than that of LM algorithm, because the model training under LM algorithm automatically stops when the generalization stops improving, which means it requires less training time but is lack of accuracy. Differently, the model training through Bayesian regularization stops according to adaptive weight minimization, leading to a higher accuracy, especially for the prediction of complex and noisy datasets.

C. PREDICTION OF ADMET PROPERTIES

To make prediction of the ADMET properties closer to reality, the ADMET properties tables of 1974 compounds are downloaded from ZINC database. For facilitating modeling,

only five ADMET properties of compounds are considered: 1) Caco2, which is used to measure the ability of compounds to be absorbed by human body; 2) CYP3A4, which is the main metabolic enzyme in human body, used to measure the metabolic stability of compounds; 3) hERG, which measures the cardiotoxicity of the compound; 4) HoB, a property can measure the proportion of drugs absorbed into the human blood circulation after it enters the human body; 5) MN, a method to detect whether the compounds have genotoxicity.

According to the empirical formula, when the input port is 20 and the output port is 1, the number of nodes in the middle layer is between 28 and 38, which is a relatively optimal value. Therefore, the numbers of intermediate nodes of the constructed neural network model are tested, and the test accuracy of the five properties of ADMET with the best training data of each node number are plotted. As Figure 7 shows, Caco2, CYP3A4, hERG, HoB, MN can be obtained the best prediction performance of 92.2%, 94.3%, 90.2%, 87.2%, and 92.2%, respectively, when the numbers of hidden nodes are 34, 33, 30, 37, 28, severally. Through appropriate hidden layer nodes, the highest prediction accuracy of the properties of each compound is achieved.

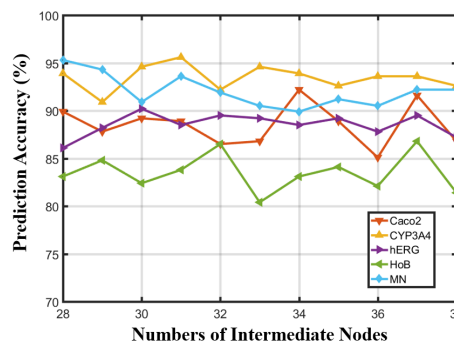


FIGURE 7. Relationship between the number of intermediate nodes and prediction accuracy.

In order to predict the ADMET performance of other unknown compounds, Cross Entropy loss function is used to judge the performance of the neural network model on the samples. It divides 1974 groups of sample data into training set, validation set and test set in the ratio of 14:3:3. The training set is used to train the network and adjust the weights of the internal transmission matrix, the function of the verification set is to find the best number of training epochs, whereas the test set acts as the final performance evaluation of the neural network model. Taking CYP3A4 as an example, the confusion matrix evaluation diagrams will be obtained as Figure 8, with the green modules indicating correct predictions, and the red modules indicating that the predictions are wrong.

Figure 8 represents that in the training data, 334 plus 1011 of the main diagonal are the sum of the correct parts of the training, accounting for 97.3%, whereas the error accounts for 2.7%. With the same method, the sum of correct parts in validation data, test data and summary data are



FIGURE 8. CYP3A4 prediction model evaluation confusion matrix. (a). Training confusion matrix. (b). Validation confusion matrix. (c). Test confusion matrix. (d). All confusion matrix.

94.9%, 94.3%, and 96.5%. That is, for the neural network model we trained, through 729 parameters, the accuracy of CYP3A4 data prediction can reach up to 94.3%. It shows that the prediction accuracy of our model is convincingly high.

The Receiver Operating Characteristic (ROC) curve is usually used to measure the performance of the model. The abscissa of the plane is False Positive Rate (FPR), and the ordinate is True Positive Rate (TPR). For a classifier, we can acquire a TPR and FPR point pair according to its performance on the test sample. Generally, the closer ROC curve is to the (0, 1) point, the better the prediction performance of the binary classification model is. The ROC curve can be calculated from the confusion matrix in Figure 8, as shown in Figure 9. The best points on ROC curve of training set, verification set, and test set are (0.054, 0.946), (0.075, 0.925), (0.05, 0.95), respectively. Boltzmann-Enhanced Discrimination of ROC (BEDROC) is a score that is more representative of compound prioritization, since it is biased towards early enrichment. The area under ROC curve is called AUC, it can be easily calculated to measure the model.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (13)$$

Another parameter to measure the quality of the model, Matthews Correlation Coefficient (MCC), is the correlation coefficient between observed and predicted binary classifications, with the values staying between $[-1, +1]$. When the coefficient equals to +1, it indicates perfect prediction,

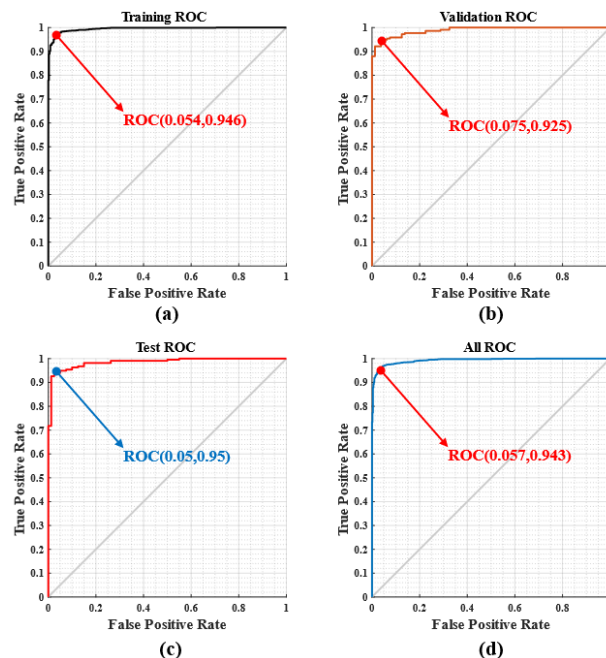


FIGURE 9. ROC curve of CYP3A4 prediction evaluation model. (a). Training ROC curve. (b). Validation ROC curve. (c). Test ROC curve. (d). All ROC curve.

0 indicates that it is no better than random prediction, and -1 refers to complete inconsistency between prediction and observation. MCC can be calculated directly from the confusion matrix using formula (13). In this formula, TP stands for true positive, TN represents true negative, FP acts for false positive and FN means false negative.

As shown in Table 1, five general parameters are used to measure our model. The BEDROC values predicted by several widely used methods such as NB, SVM, LR, DNN_PCM are respectively 0.79 ± 0.08 , 0.88 ± 0.05 , 0.88 ± 0.06 , and 0.93 ± 0.03 . The BEDROC value predicted by the three-layer neural network model under the Bayesian regularization algorithm we proposed can reach 0.95 ± 0.03 , which is 2-3% higher than the highest value of the previous models. The MCC values predicted by NB, SVM, LR, DNN_PCM are 0.41 ± 0.03 , 0.56 ± 0.07 , 0.51 ± 0.06 , and 0.53 ± 0.07 , sequentially. The MCC value of our model is 0.93 ± 0.03 , which is 37-40% higher than the highest MCC value of the previous models. The AUC values predicted by NB, SVM, LR are correspondingly 0.7875, 0.8126, 0.8402, whereas our AUC value is 0.9752, which is the highest among them. The data comparisons indicate that the prediction performance of our proposed model is extraordinary.

After processing 1974 samples, the remaining 543 samples with high correlation are selected. The data of the first 20 samples with better biological activity to inhibit ER α are then processed, as shown in Figure 10. Taking the molecular Lipoaffinity Index as an example, when the molecular descriptor value stays from 10 to 12, the ADMET properties expresses with the highest frequency, about 28 times. For

TABLE 1. Model performance comparison table.

Method and Algorithm	NB	SVM	LR	DNN_PCM	Our Work
BEDROC ^[12]	0.79±0.08	0.88±0.05	0.88 ± 0.06	0.93±0.03	0.95±0.03
MCC ^[13]	0.41±0.03	0.56±0.07	0.51 ± 0.06	0.53±0.07	0.93±0.03
AUC ^[14]	0.7875	0.8126	0.8402	/	0.9752
R	/	/	/	/	0.82205
MSE	/	/	/	/	0.40184

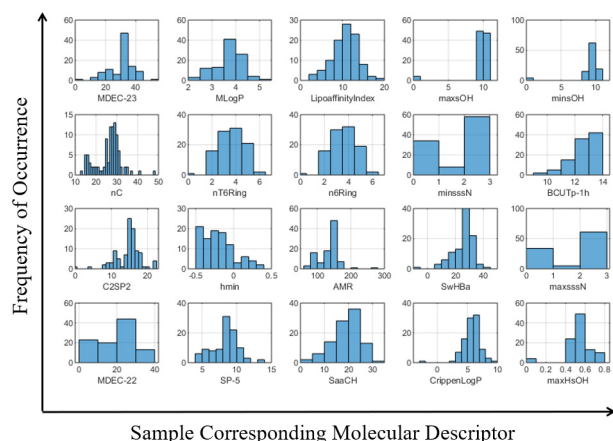


FIGURE 10. Histogram of occurrence frequency of molecular descriptors corresponding to samples.

molecular n6Ring, when the molecular descriptor value is between 2.5 to 4.5, the frequency of ADMET properties reaches the highest, about 32 times. 20 samples are listed in Figure 10, who are compounds screened by our proposed model that not only have high anti-ER α activity, but also show excellent ADMET properties for multiple times in a specific interval. The result shows that the model we designed has successfully achieved the purpose of screening effective anticancer drugs with little side effects. In recent years, deep learning has also been used to detect COVID-19 images [32], [33]. It is believed that in the near future, deep learning will play an important role in drug screening and virus detection, and bring new breakthroughs to the development of medicine.

IV. CONCLUSION

For future drug screening, the contribution of computers can greatly reduce meaningless steps and provide strong support for drug discovery. In dealing with the massive number of parameters that need to be considered in model training, this paper presents a method that takes into account both the burden of computing power and prediction accuracy. An efficient quantitative prediction model (with AUC of 0.9752, MCC of 0.93) of compounds bioactivity has been constructed. During the training, Pearson correlation analysis, Levenberg Marquardt algorithm and Bayesian regularization algorithm were used to effectively train and analyze the data. Moreover, the ADMET properties were also considered, using cross-entropy to calculate the loss of neural network model and

supervise learning. Our training data sets are more extensive and the prediction accuracy of our proposed model is higher, with the values of AUC and MCC much higher than other methods. This research can improve the efficiency of screening the compounds who can inhibit the ER α bioactive, providing new ideas for the treatment of breast cancer.

REFERENCES

- [1] X. Feng, L. Song, S. Wang, H. Song, H. Chen, Y. Liu, C. Lou, J. Zhao, Q. Liu, Y. Liu, and R. Zhao, "Accurate prediction of neoadjuvant chemotherapy pathological complete remission (pCR) for the four subtypes of breast cancer," *IEEE Access*, vol. 7, pp. 134697–134706, 2019, doi: 10.1109/ACCESS.2019.2941543.
- [2] E. Kontopodis, M. Venianaki, G. C. Manikis, K. Nikiforaki, O. Salvetti, E. Papadaki, G. Z. Papadakis, A. H. Karantanas, and K. Marias, "Investigating the role of model-based and model-free imaging biomarkers as early predictors of neoadjuvant breast cancer therapy outcome," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 5, pp. 1834–1843, Sep. 2019, doi: 10.1109/JBHI.2019.2895459.
- [3] J. C. Nava-Ramirez, S. E. Santana-Krimskaya, M. A. Franco-Molina, A. S. Ortega-Villarreal, I. Lopez, D. J. Michaelis, and E. Hernandez-Fernandez, "Synthesis of α , β -unsaturated benzotriazolyl-1,3,4-oxadiazole derivatives: Anticancer activity, cytotoxicity, and cell imaging," *IEEE Trans. Nanobiosci.*, vol. 21, no. 1, pp. 125–134, Jan. 2022, doi: 10.1109/TNB.2021.3100888.
- [4] J. Y. Pae, R. V. Nair, P. Padmanabhan, G. Radhakrishnan, B. Gulyás, and M. V. Matham, "Gold nano-urchins enhanced surface plasmon resonance (SPR) BIOSENSORS for the detection of estrogen receptor alpha (ER α)," *IEEE J. Sel. Topics Quantum Electron.*, vol. 27, no. 5, pp. 1–6, Sep. 2021, doi: 10.1109/JSTQE.2021.3069453.
- [5] B. Fu, P. Liu, J. Lin, L. Deng, K. Hu, and H. Zheng, "Predicting invasive disease-free survival for early stage breast cancer patients using follow-up clinical data," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 7, pp. 2053–2064, Jul. 2019, doi: 10.1109/TBME.2018.2882867.
- [6] P. Liu, B. Fu, S. X. Yang, L. Deng, X. Zhong, and H. Zheng, "Optimizing survival analysis of XGBoost for ties to predict disease progression of breast cancer," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 1, pp. 148–160, Jan. 2021, doi: 10.1109/TBME.2020.2993278.
- [7] S. Maldonado, J. Merigo, and J. Miranda, "IOWA-SVM: A density-based weighting strategy for SVM classification via OWA operators," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 9, pp. 2143–2150, Sep. 2020, doi: 10.1109/TFUZZ.2019.2930942.
- [8] J. Liang, Z. Qin, J. Ni, X. Lin, and X. Shen, "Practical and secure SVM classification for cloud-based remote clinical decision services," *IEEE Trans. Comput.*, vol. 70, no. 10, pp. 1612–1625, Oct. 2021, doi: 10.1109/TC.2020.3020545.
- [9] G. Alexandrou, N. Moser, K.-T. Mantikas, J. Rodriguez-Manzano, S. Ali, R. C. Coombes, J. Shaw, P. Georgiou, C. Toumazou, and M. Kalofonou, "Detection of multiple breast cancer ESR1 mutations on an ISFET based lab-on-chip platform," *IEEE Trans. Biomed. Circuits Syst.*, vol. 15, no. 3, pp. 380–389, Jun. 2021, doi: 10.1109/TBCAS.2021.3094464.
- [10] J. Cao, G. Lv, C. Chang, and H. Li, "A feature selection based serial SVM ensemble classifier," *IEEE Access*, vol. 7, pp. 144516–144523, 2019, doi: 10.1109/ACCESS.2019.2917310.
- [11] W. Kudisthalert and K. Pasupa, "A coefficient comparison of weighted similarity extreme learning machine for drug screening," in *Proc. 8th Int. Conf. Knowl. Smart Technol. (KST)*, Feb. 2016, pp. 43–48, doi: 10.1109/KST.2016.7440525.

- [12] K. A. Carpenter, D. S. Cohen, J. T. Jarrell, and X. Huang, "Deep learning and virtual drug screening," *Future Medicinal Chem.*, vol. 10, no. 21, pp. 2557–2567, Nov. 2018, doi: [10.4155/fmc-2018-0314](https://doi.org/10.4155/fmc-2018-0314).
- [13] E. B. Lenselink, N. ten Dijke, B. Bongers, G. Papadatos, H. W. T. van Vlijmen, W. Kowalczyk, A. P. IJzerman, and G. J. P. van Westen, "Beyond the hype: Deep neural networks outperform established methods using a ChEMBL bioactivity benchmark set," *J. Cheminformatics*, vol. 9, no. 1, p. 45, Aug. 2017, doi: [10.1186/s13321-017-0232-0](https://doi.org/10.1186/s13321-017-0232-0).
- [14] P. Ding, R. Yin, J. Luo, and C.-K. Kwok, "Ensemble prediction of synergistic drug combinations incorporating biological, chemical, pharmacological, and network knowledge," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 3, pp. 1336–1345, May 2019, doi: [10.1109/JBHI.2018.2852274](https://doi.org/10.1109/JBHI.2018.2852274).
- [15] G. Urban, "Deep learning for drug discovery and cancer research: Automated analysis of vascularization images," *IEEE/ACM Trans. Comput. Bio. Bioinf.*, vol. 16, no. 3, pp. 1029–1035, May/Jun. 2019, doi: [10.1109/TCBB.2018.2841396](https://doi.org/10.1109/TCBB.2018.2841396).
- [16] A. Pezeshk, S. Hamidian, N. Petrick, and B. Sahiner, "3-D convolutional neural networks for automatic detection of pulmonary nodules in chest CT," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 5, pp. 2080–2090, Sep. 2019, doi: [10.1109/JBHI.2018.2879449](https://doi.org/10.1109/JBHI.2018.2879449).
- [17] A. Ion, S. Gosav, and M. Praisler, "Artificial neural networks designed to identify NBOMe hallucinogens based on the most sensitive molecular descriptors," in *Proc. 6th Int. Symp. Electr. Electron. Eng. (ISEEE)*, Oct. 2019, pp. 1–6, doi: [10.1109/ISEEE48094.2019.9136101](https://doi.org/10.1109/ISEEE48094.2019.9136101).
- [18] D. Zhou, S. Peng, D.-Q. Wei, W. Zhong, Y. Dou, and X. Xie, "LUNAR: Drug screening for novel coronavirus based on representation learning graph convolutional network," *IEEE/ACM Trans. Comput. Bio. Bioinf.*, vol. 18, no. 4, pp. 1290–1298, Jul. 2021, doi: [10.1109/TCBB.2021.3085972](https://doi.org/10.1109/TCBB.2021.3085972).
- [19] J.-L. Lin, W.-L. Kuo, Y.-H. Huang, T.-L. Jong, A.-L. Hsu, and W.-H. Hsu, "Using convolutional neural networks to measure the physiological age of *Caenorhabditis elegans*," *IEEE/ACM Trans. Comput. Bio. Bioinf.*, vol. 18, no. 6, pp. 2724–2732, Nov. 2021, doi: [10.1109/TCBB.2020.2971992](https://doi.org/10.1109/TCBB.2020.2971992).
- [20] A. Morro, V. Canals, A. Oliver, M. L. Alomar, F. Galan-Prado, P. J. Ballester, and J. L. Rossello, "A stochastic spiking neural network for virtual screening," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 4, pp. 1371–1375, Apr. 2018, doi: [10.1109/TNNLS.2017.2657601](https://doi.org/10.1109/TNNLS.2017.2657601).
- [21] Y. Lai, G. Li, D. Wu, W. Lian, C. Li, J. Tian, X. Ma, H. Chen, W. Xu, J. Wei, Y. Zhang, and G. Jiang, "2019 novel coronavirus-infected pneumonia on CT: A feasibility study of few-shot learning for computerized diagnosis of emergency diseases," *IEEE Access*, vol. 8, pp. 194158–194165, 2020, doi: [10.1109/ACCESS.2020.3033069](https://doi.org/10.1109/ACCESS.2020.3033069).
- [22] L. Kong and H. Nian, "Fault detection and location method for mesh-type DC microgrid using Pearson correlation coefficient," *IEEE Trans. Power Del.*, vol. 36, no. 3, pp. 1428–1439, Jun. 2021, doi: [10.1109/TPWRD.2020.3008924](https://doi.org/10.1109/TPWRD.2020.3008924).
- [23] H. Zhu, X. You, and S. Liu, "Multiple ant colony optimization based on Pearson correlation coefficient," *IEEE Access*, vol. 7, pp. 61628–61638, 2019, doi: [10.1109/ACCESS.2019.2915673](https://doi.org/10.1109/ACCESS.2019.2915673).
- [24] J. Liu, A. R. Sternberg, S. Ghiasvand, and Y. Berdichevsky, "Epilepsy-on-a-chip system for antiepileptic drug discovery," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 5, pp. 1231–1241, May 2019, doi: [10.1109/TBME.2018.2871415](https://doi.org/10.1109/TBME.2018.2871415).
- [25] N. Meng, E. Y. Lam, K. K. Tsia, and H. K.-H. So, "Large-scale multi-class image-based cell classification with deep learning," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 5, pp. 2091–2098, Sep. 2019, doi: [10.1109/JBHI.2018.2878878](https://doi.org/10.1109/JBHI.2018.2878878).
- [26] G. Huang and Q. Yan, "Optimizing features quality: A normalized covariance fusion framework for skeleton action recognition," *IEEE Access*, vol. 8, pp. 211869–211881, 2020, doi: [10.1109/ACCESS.2020.3037238](https://doi.org/10.1109/ACCESS.2020.3037238).
- [27] M. Li, Y. Wang, R. Zheng, X. Shi, Y. Li, F.-X. Wu, and J. Wang, "DeepDSC: A deep learning method to predict drug sensitivity of cancer cell lines," *IEEE/ACM Trans. Comput. Bio. Bioinf.*, vol. 18, no. 2, pp. 575–582, Mar. 2021, doi: [10.1109/TCBB.2019.2919581](https://doi.org/10.1109/TCBB.2019.2919581).
- [28] M. Imani and U. M. Braga-Neto, "Control of gene regulatory networks using Bayesian inverse reinforcement learning," *IEEE/ACM Trans. Comput. Bio. Bioinf.*, vol. 16, no. 4, pp. 1250–1261, Jul. 2019, doi: [10.1109/TCBB.2018.2830357](https://doi.org/10.1109/TCBB.2018.2830357).
- [29] D. Gaglione, P. Braca, L. M. Millefiori, G. Soldi, N. Forti, S. Marano, P. K. Willett, and K. R. Pattipati, "Adaptive Bayesian learning and forecasting of epidemic evolution—Data analysis of the COVID-19 outbreak," *IEEE Access*, vol. 8, pp. 175244–175264, 2020, doi: [10.1109/ACCESS.2020.3019922](https://doi.org/10.1109/ACCESS.2020.3019922).
- [30] H. Chen, D. A. K. Maduranga, P. A. Mundra, and J. Zheng, "Bayesian data fusion of gene expression and histone modification profiles for inference of gene regulatory network," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 17, no. 2, pp. 516–525, Mar. 2020, doi: [10.1109/TCBB.2018.2869590](https://doi.org/10.1109/TCBB.2018.2869590).
- [31] M. Wu, L. Zhong, B. Xu, and N. Xiong, "A consensus-based diffusion Levenberg-Marquardt method for collaborative localization with extension to distributed optimization," *IEEE Access*, vol. 8, pp. 215649–215660, 2020, doi: [10.1109/ACCESS.2020.3041491](https://doi.org/10.1109/ACCESS.2020.3041491).
- [32] M. R. H. Mondal, S. Bharati, and P. Podder, "CO-IRv2: Optimized InceptionResNetV2 for COVID-19 detection from chest CT images," *PLoS ONE*, vol. 16, p. 10, Oct. 2021, doi: [10.1371/journal.pone.0259179](https://doi.org/10.1371/journal.pone.0259179).
- [33] S. Bharati, P. Podder, M. R. H. Mondal, and V. B. S. Prasath, "Medical imaging with deep learning for COVID-19 diagnosis: A comprehensive review," *Int. J. Comput. Inf. Syst. Ind. Manage. Appl.*, vol. 13, pp. 2150–2158, Jul. 2021, doi: [10.48550/arXiv.2107.09602](https://doi.org/10.48550/arXiv.2107.09602).



ZHANGMIAOGE LIU was born in 1998. She received the Bachelor of Engineering degree in electronic science and technology from the University of Shanghai for Science and Technology, Shanghai, China, where she is currently pursuing the Master of Engineering degree in optical engineering. Her research interests include neural networks, drug screening, and thermoelectric cooling.



NING WANG (Member, IEEE) was born in 1984. He received the M.S. and Ph.D. degrees in microelectronics from Xidian University, Xi'an, China, in 2009 and 2012, respectively. In 2012, he joined China North Industries Group Corporation, General Electronics Group, Suzhou, China, as a Research and Development Engineer and the Project Leader. From 2013 to 2015, he was a Postdoctoral Researcher with Shanghai Jiao Tong University, Shanghai. Since 2015, he has been a

Faculty Member with the University of Shanghai for Science and Technology, Shanghai, where he is currently an Associate Professor. His current research interests include neural networks, data screening, thermoelectric energy conversion, high-speed interconnect technology, and waveguide in terahertz.



CHEN ZHANG was born in 1999. He received the Bachelor of Engineering degree in electronic science and technology from the University of Shanghai for Science and Technology, Shanghai, China, where he is currently pursuing the Master of Engineering degree in optical engineering. His research interests include drug screening, neural networks, and thermoelectric energy conversion.



ZHOUXIAO LIU was born in 1996. He received the bachelor's and master's degrees in engineering from the University of Shanghai for Science and Technology, Shanghai, China, in 2018 and 2021, respectively. He is currently engaged in semiconductor industry. His research interests include neural networks, data screening, thermoelectric refrigeration, and energy collection.