A Robust Regression-Based Modeling to Predict Antiplasmodial Activity of Thiazolyl-Pyrimidine Hybrid Derivatives against *Plasmodium falciparum*

Kevin S. Umoette, Charles O. Nnadi, Wilfred O. Obonga

Department of Pharmaceutical and Medicinal Chemistry, University of Nigeria Nsukka,, Nigeria.

INTRODUCTION

Resistance to current available antimalarial drugs poses a severe threat to the elimination of malaria, which results in a sharp rise in the number of deaths each year, as well as increased medical expenses and lost productivity. Drug development takes approximately 14 years from necessary pre-clinical testing to regulatory approval due to the challenges that arise during the process. Several heterocycles have demonstrated antiplasmodial activity. Thiazolyl-pyrimidine hybrid plays significant roles in the biological activities and SAR of thiazolylpyrimidines (Tzpd), thiazolopyrimidines and thienopyrimidines due the combination of the thiazole and pyrimidine pharmacophores (**Fig. 1**). .

RESULTS

Of the 145 features calculated for the 43 Tzpd, 6 molecular features: FCASA-, MNDO_LUMO, E_str, vsurf_HB1, vsurf_G and vsurf_DD12 ($p < 0.05$; VIF < 5) were found to significantly influence the antiplasmodial activity. Five-fold cross-validation performance scores of MLR, kNN, SVR, and RFR showed that the performance metrics of MLR (MSE = 0.1453; R^2 = 0.68; MAE = 0.290; RMSE = 0.381; pIC_{50}(predicted) = 8.06 - 0.45vsurf_G + 0.37FCASA– - 0.42MNDO_LUMO – 0.20E_str + 0.30vsurf_HB1 – 0.38vsurf_DD12) outperformed other models..

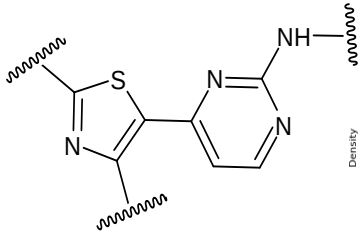


Fig. 1. Pharmacophore of thiazolyl-pyrimidine hybrid derivatives

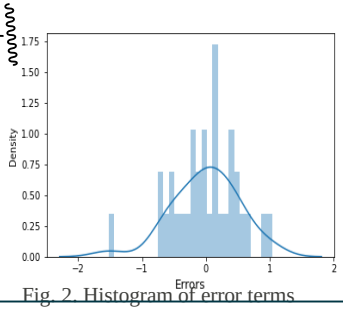
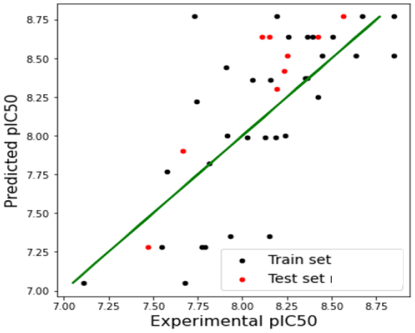


Fig. 2. Histogram of error terms

Residue analysis of the error terms were checked to ascertain their normal distribution. Since the error terms are normally distributed (**Fig. 2**), the model can be used to make predictions on the test dataset.

METHODS

The study developed regression-based models for the prediction of antiplasmodial activity of 43 Tzpd hybrid obtained from the ChEMBL database. The molecular descriptors (145 features) were scaled down to 6 using the recursive feature elimination. The X- and Y-matrix were split into 34 train and 9 test sets using a split ratio of 0.20. Regression models were built using scikit-learn algorithms: multiple linear regressor (MLR), k-Nearest Neighbours (kNN), Support Vector Regressor (SVR) and Random Forest Regressor (RFR)) to predict the pIC_{50} of the test set. The models were evaluated using R^2, mean squared error (MSE), mean absolute error (MAE), root mean squared error (RMSE), p-values, F-statistic, and variance inflation factor (VIF)



To prove further confidence in our predicted pIC_{50} values, the predicted pIC_{50} scores were plotted against the experimental pIC_{50} scores for both the train set and the test set, using MLR model (**Fig 3**). The R^2 indicates how closely the data resemble the regression line and how well the data fit the regression line.

Table 1.. Model evaluation and comparison

Algorithms	kNN	SVR	RFR	MLR
Test MSE	0.00	0.053	0.069	0.1453
5-fold CV	0.59±0.41	0.67±0.45	0.75±0.29	0.091±0.010
Test R^2	1.00	0.61	0.36	0.680
5-fold CV	0.36±0.46	0.63±0.62	0.59±2.21	0.745±0.281
Test MAE	0.00	0.174	0.209	0.290
5-fold CV	0.55±0.18	0.58±0.20	0.60±0.60	0.270±0.101
Test RMSE	0.00	0.230	0.262	0.381
5-fold CV	0.72±0.27	0.77±0.27	0.84±0.18	0.302±0.021

CV = Cross- validation

The correlations of the predicted and experimental solubility values are shown in Table 1. The R^2 indicates how closely the data resemble the regression line and how well the data fit the regression line. For the MLR model, R^2 values for the train and test sets are 0.68. The R^2 values for the train and test sets when the SVR model was used were 0.61.

CONCLUSION

❖The study developed predictive models and provided insights into the chemical features necessary for the optimization of thiazolyl-pyrimidine to enhance antiplasmodial activity.

REFERENCES

1. Afuwape *et al*., Comput Standards Interfaces. **2023**
2. Lee *et al*., *ACS Omega*. **2022**
3. Oguike *et al*., Mol diver. **2022**