Stacked Ensemble for Bioactive Molecule Prediction

OLUTOMILAYO OLAYEMI PETINRIN[1,2] AND FAISAL SAEED[1,3]

[1]Information Systems Department, Faculty of Computing, Universiti Teknologi Malaysia Johor Bahru 81310, Malaysia
[2]Department of Computer Science, City University of Hong Kong, Hong Kong
[3]College of Computer Science and Engineering, Taibah University, Medina 42353, Saudi Arabia

Corresponding author: Faisal Saeed (fsaeed@taibahu.edu.sa)

ABSTRACT Bioactive molecular compounds are essential for drug discovery. The biological activity of these compounds needs to be predicted as this is used to determine the drug-target ability. As ineffective drugs are discarded after production, leading to resource and time wastage, it is important to predict bioactive molecules with models having high predictive performance. This study utilizes the stacked ensemble which uses the prediction of multiple base classifiers as features, used to train a meta classifier which makes the final prediction. Using three datasets DS1, DS2, and DS3 gotten from MDL Drug Data Report (MDDR) database, the performance of stacked ensemble was compared to three other ensembles: adaboost, bagging, and vote ensemble, based on different evaluation criteria and also a statistical method, Kendall's W test. The accuracy of Stacked ensemble ranged from 96.7002%, 98.2260% and 94.9007% for the three datasets respectively, although Vote had the best accuracy using dataset DS2 which consist of structurally homogeneous bioactive molecules. Also, using Kendall's W test to rank the ensembles, Stacked ensemble was ranked best with datasets DS1 and DS3, with both having a mean average of 4.00 and an overall level of agreement, W, of 0.986 and 1.000 respectively. Using dataset DS2, it was ranked after Vote and Adaboost with mean average of 2.33 and an overall level of agreement, W of 0.857. Stacked ensemble is recommended for the prediction of heterogeneous bioactive molecules during drug discovery and can also be implemented in other research areas.

INDEX TERMS Bioactive molecule prediction, chemoinformatics, drug discovery, ensemble, stacked ensemble.

I. INTRODUCTION

Bioactive molecular compounds are substances with positive effects on living organisms. They have toxicological and pharmacological effects on humans and animals. These compounds can either be found in plants, fruits, nuts, or synthesized. Various compounds exist in a plant, including bioactive compounds. These may be present in any part of the plant and they need to be extracted for further use [1]. Some bioactive compounds can also be found in marine ecosystem such as marine sponges [2], [3]. Nopal cactus or prickly pear fruit is also a plant which has been used in both traditional medicine, and for its bioactive compounds properties [4]. Examples of bioactive compounds are flavonoids, carotenoids, and polyphenols.

The associate editor coordinating the review of this manuscript and approving it for publication was Mahmoud Barhamgi.

Bioactive molecular compounds are of high importance in the drug discovery process, and it is also important that the model used in predicting these bioactive molecules are accurate enough to avoid unwanted error which leads to inability of the drug to meet the condition for which it was developed. The introduction of ensembles to prediction has allowed models to combine the performance ability of more than one classifier to improve the performance of the overall model. Ensembles can be likened to the productivity which emanates as a result of teamwork compared to a lone work.

II. RELATED WORK

The combination of two or more machine learning methods have shown great prospect at outperforming that of single method. Application of multiple classifiers [5]–[7], otherwise known as ensemble or wisdom of crowds has resulted in better accuracy and performance compared to single classifiers.

As no algorithm can be said to be the most accurate, and since different classifiers are associated with different algorithms, representations, parameters, problems, and training sets, problems associated with each classifier used is shared by the ensemble.

Stacked ensemble, also known as stacked generalization is an ensemble method proposed and introduced by Wolpert [8] in 1992. It has been implemented in some areas of science for prediction, but its implementation is yet to be found in bioactive molecule prediction. He *et al.* [9] utilized stacked generalization in extracting literatures which are related to drug-drug interaction from the numerous literatures available. Although the f-score achieved from the study could be considered low, it was a remarkable improvement compared to previous f-score from other systems. Also, Anifowose *et al.*, [10] implemented an homogeneous stacked ensemble using Support Vector Machine (SVM) in the prediction of petroleum reservoir characterization. Using bagging method and random forest (RF) with support vector machine also, it was reported that in most cases, stacked ensemble with support vector machine had better performance than the other methods and that better result was achieved compared to using support vector machine only. Gupta and Thakkar [11] also made a comparative study on three metaheuristic algorithms which has been used in optimizing stacked ensemble. These algorithms are: genetic algorithm, particle swarm optimization, and ant colony optimization, although it was reported that particle swarm optimization can provide better results. In a previous study [12], artificial bee colony was implemented with 10 datasets and its effectiveness in the selection of metaclassifier and base classifier was shown. Stacked ensemble has also been used in analyzing teenage distress [13], and brain related study [14].

The effectiveness of stacked ensemble can also be introduced to bioactive molecule prediction. This study which implements stacked ensemble will be compared against some other ensembles which has been previously used in bioactive molecule prediction.

A. BAGGING

Bootstrap Aggregating, also known as Bagging, is an ensemble method which is capable of handling both classification and regression. It is a parallel ensemble which reduces both variance and overfitting but does not reduce bias.

Bagging method works by creating m number of bags, and each bag contains a subset of the original data. If the original dataset consists of n different instances, n^1 of the original data is taken with replacement into each bag. Replacement implies that the same instance or data point can be chosen more than once since it is a random selection, and an instance can be several other bags. Therefore, all m number of bags contains n^1 different instances which are all chosen with replacement. According to the rule of thumb, n^1 is usually about 60% of n, that is, each bag contains about 60% of the original data. In the case of this study, n^1 was 100% of n since it gave a better result compared to 60% and, the number of bags m,

created is 10. Each data in each bag is used to train the model using a classifer (in the case of this study, random forest was used as the base classifier). Therefore, the output X, which is the result from each model is used to determine the overall result Y using either the mean of all the output for regression, or the mode of all the output in the case of classification.

Bagging has also been used for intrusion detection systems [15], classification of arrhythmia [16], forecasting the flow of traffic [17], and also in conjunction with boosting [18], where the two ensembles have shown capacity of having better performance than single classifiers

B. ADABOOST

Adaboost, developed by Yoar Freund and Robert Schapire [19], [20], is the short form of Adaptive Boosting, and it is a meta-algorithm which is the commonly known and used form of Boosting. Boosting is a fairly simple variation on bagging ensemble. It tries to improve learners by focusing on areas where the system is not performing well. Adaboost is a sequential ensemble method. It builds multiple models of the same classifier each of which learns to fix the prediction errors of the prior model in the chain. The classifier's output is combined into a weighted sum and this denotes the final output of the boosted classifiers. This boosting method is called adaptive because subsequent weak learners are tweaked in favor of misclassified instances.

Adaboost weighs instances in the dataset by how easy or difficult they are to classify, and this allows the algorithm to pay more attention to them when constructing subsequent models. The wrongly classified instances get high weight and each subsequent boosting learns a new classifier on the weighted dataset. The classifiers are then weighed to combine them into a single powerful classifier. Those classifiers with low training error rate have high weight. The process is halted with cross validation. Adaboost has shown better performance compared to Support Vector Machine, Artificial Neural Network, and K-Nearest Neighbor in the analysis of the Structure-Activity relationship of phenol [21], and also improved classification time and accuracy in the classification of pecan defects [22].

C. VOTE ENSEMBLE

Vote ensemble can simply and effectively used for problems associated with classification. This method works on multiclassifiers to reduce misclassification. It has demonstrated significant performance in enhancing the accuracy of predictions. It has been applied to a number of areas and impressive results have been achieved. According to Yan *et al.* [23], voting based method has shown effectiveness in managing incomplete data without making presumptions about missing values. Vote ensemble utilizes several combination algorithms to makes it predictions. These combination rules include: Average Probabilities, Minimum Probabilities, Maximum Probabilities, Product of Probabilities, and Majority Voting [24], . It creates series of classifiers and then predicts

based on either the mode or mean of the base classifiers. Majority Voting has been used majorly for prediction as the output of the ensemble or classification is the label with the highest number of votes from the base classifiers [24]–[28]. It can also be weighted [30], that is, assigning more weight on classifiers which are more likely correct [31].

Choosing classification methods with uncorrelated predictions is essential. It is known that Bayesian, Decision Tree (DT) and K-Nearest Neighbor (K-NN) should be selected in order to have classifiers with diverse predictions. It has been used for clustering in chemoinformatics [32], [33], and detection of mislabelled data in bioinformatics [34]. This study is carried out using majority voting as the combination rule.

III. EXPERIMENTAL DESIGN
SVM, DT, K-NN, and RF were used as base classifiers for Stacked ensemble and Voting ensemble. Random Forest, due to its high performance in classification [35]–[41] was used as the meta-classifier in the Stacked ensemble to enable better prediction, and it was also used as the base classifier for Adaboost, and Bagging ensemble.

Waikato Environment for Knowledge Analysis (WEKA), a software written in java, for machine learning was used as the cross-platform with which analysis was made.

A. DATASETS
The study was implemented using three datasets obtained from the MDL Drug Data Report (MDDR) which has already been converted to Pipeline Pilot's ECFC_4 fingerprints and folded into a size of 1024 element fingerprints. These datasets have been used extensively in previous chemoinformatics researches such as virtual screening [42]–[44], and prediction of bioactive molecule [45], [46].

The bioactivity of the molecules was the class on which prediction was made. DS1 consist of 8294 structurally heterogeneous and homogeneous bioactive molecules and 11 classes. DS2 consist of 5083 homogeneous bioactive molecules with 10 classes and DS3 consist of 8560 heterogenous bioactive molecules with 10 classes also. Each dataset was divided into training and testing data with 70:30 ratio respectively. The three datasets are described and used in [42]–[44].

B. STACKED ENSEMBLE MECHANISM
Stacked ensemble is in the form of a two-level procedure. It deals with multiple classifiers and, also, a metaclassifier. On the first level, several base classifiers are introduced to make predictions. This can either be homogeneous, that is, same type classifier, or heterogeneous, several types of classifier. In this paper, an heterogenous set of classifiers consisting of Decision Tree, Support Vector Machine, Random Forest, and K-Nearest Neighbor were used. On the second level, the outputs from the base classifiers are introduced to the metaclassifier, Random Forest in this paper, which is trained based on these outputs and learns the best way to correct the

errors from the base classifiers before generating the final prediction.

It is imperative that the base classifiers produce different predictions which are uncorrelated. Better performance is achieved using different algorithms. Basically, stacked ensemble follows the following procedure:

1) Several base classifiers are trained using the training data. It is better to select different base classifiers with different algorithms
2) A meta-classifier is then trained based on the predictions from the base classifiers. This meta-classifier learns when the base classifiers are right or wrong.
3) Final prediction is made by the meta-classifier using the base classifiers prediction for the testing data.

The algorithm for stacking is shown in Algorithm 1 and the mechanism is depicted in Figure 1.

Algorithm 1 Stacking

1. **Input**: Training data
$$D = \{a_i, b_i\}_{i=1}^m (x_i \in R^n, y_i \in Y)$$
2. Step 1: Learn base-level classifiers
3. **for** t ← 1 to T **do**
4.   Learn a base classifier S_t based on D
5. **end for**
6. Step 2: Construct new data sets from D
7. **for** i ← to m **do**
8.   Construct a new dataset that contains $D_s = \{x_i^1, y_i\}$, where $x_i^1 = \{s_1(x_i), s_2(x_i), \ldots, s_T(x_i)\}$
9. **end for**
10. Step 3: Learn a meta classifier
11. Learn an ensemble classifier S based on the new constructed dataset D_s
12. **return** $S(x) = s^1(s_1(x), s_2(x), \ldots, s_T(x))$
13. **Output**: Ensemble Classifier S
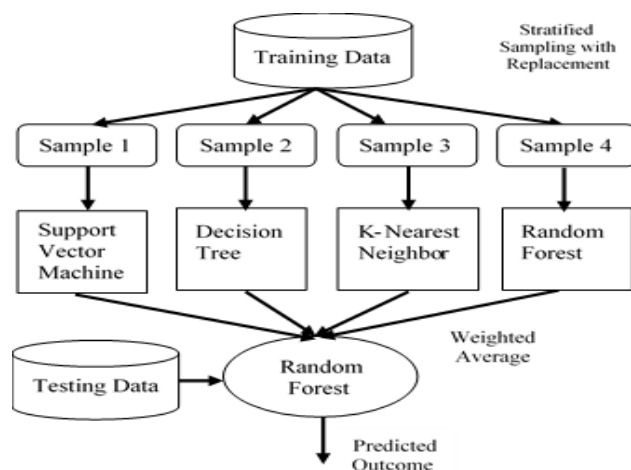

FIGURE 1. Stacked ensemble mechanism.

IV. RESULTS AND DISCUSSION

The performance of three ensembles - Voting, Adaboost, and Bagging were compared to the performance of the stacking ensemble based on the accuracy, sensitivity, specificity, precision, recall, and f-measure. For ensembles requiring more than one base classifier, such as Stacked, and Vote ensemble, Support Vector Machine, k-Nearest Neighbor, Decision Tree, and Random Forest were used as the base classifier, while Random Forest being a classifier with high performance was used for Adaboost and Bagging ensembles and also as the meta-classifier for stacked ensemble.

The performance of each ensemble based on the generally used evaluation criteria are shown in Table 1, 2, and 3 for datasets DS1, DS2, and DS3 respectively. Also, the ensembles were ranked using a statistical method, Kendall's Coefficient of Concordance, to ascertain which ensemble had the best performance, and the result is shown in Table 4. The highest results are bolded for emphasis.

TABLE 1. Performance evaluation of ensembles for dataset DS1.

Ensemble	Accuracy	Sensitivity	Specificity	Precision	Recall	F-measure
Stacked	**96.7002**	**0.967**	**0.997**	**0.968**	**0.967**	**0.967**
Vote	96.499	0.965	0.996	0.965	0.965	0.965
Adaboost	95.6137	0.956	0.995	0.957	0.956	0.956
Bagging	95.8551	0.959	0.995	0.959	0.959	0.958

TABLE 2. Performance evaluation of ensembles for dataset DS2.

Ensemble	Accuracy	Sensitivity	Specificity	Precision	Recall	F-Measure
Stacked	98.2260	0.982	**0.996**	0.982	0.982	0.982
Vote	**98.5545**	**0.986**	**0.996**	**0.986**	**0.986**	**0.985**
Adaboost	98.2917	0.983	0.995	0.983	0.983	0.982
Bagging	98.0289	0.980	0.995	0.980	0.980	0.979

Table 1 and 3 shows Stacked ensemble having the best overall performance based on all evaluation criteria examined. It is noted that the datasets used to generate these results contain both heterogenous and homogeneous bioactive molecules for DS1 and heterogenous bioactive molecules for DS3.

The performance of the Stacked ensemble was lower than the performance of Vote ensemble and Adaboost ensemble

TABLE 3. Performance evaluation of ensembles for dataset DS3.

Ensemble	Accuracy	Sensitivity	Specificity	Precision	Recall	F-Measure
Stacked	**94.9007**	**0.949**	**0.992**	**0.949**	**0.949**	**0.949**
Vote	94.7840	0.948	0.991	0.948	0.948	0.948
Adaboost	93.4216	0.934	0.986	0.934	0.934	0.934
Bagging	93.8887	0.939	0.988	0.940	0.939	0.939

TABLE 4. Statistical ranking of ensembles.

Dataset	Ensemble	Mean Average	Kendall's W
DS1	Stacked	**4.00**	0.986
	Vote	3.00	
	Bagging	1.92	
	Adaboost	1.08	
DS2	Stacked	2.33	0.857
	Vote	**3.92**	
	Bagging	1.08	
	Adaboost	2.67	
DS3	Stacked	**4.00**	1.000
	Vote	3.00	
	Bagging	2.00	
	Adaboost	1.00	

using the DS2 dataset as shown in Table 2. This dataset contains homogeneous molecules. Although, the comparison of the performance from the three datasets shows that DS2 generated ensembles with high prediction performance. It should be noted that the number of instances is lower compared to the instances in DS1 and DS3.

Using Kendall's coefficient of concordance in ranking the ensembles with the equivalent level of agreement between the raters for each dataset, the following ranking shown in Table 4 was obtained. For dataset DS1, the ranking was of the form, Stacked > Vote > Bagging > Adaboost, with level of agreement, W, of 0.986 and asymptotic significance, P, of 0.000. Dataset DS2 had the ranking in the form, Vote > Adaboost > Stacked > Bagging, with level of agreement, W, of 0.857 and asymptotic significance, P, of 0.001. Also, dataset DS3 had the ranking of the ensembles in the form, Stacked > Vote > Bagging > Adaboost, with level of agreement, W, of 1.000 and asymptotic significance, P, of 0.000.

As shown in Table 4, the Kendall's W, that is, level of agreement, for dataset DS2 is considerably low compared to that of DS1 and DS3. The difference between the mean average

of the ensembles in DS2 is also not has high as that of the other datasets. This shows that the ensembles performed well in the prediction. This low difference in the mean average makes it difficult for the ensembles to be effectively ranked thus leading to an agreement level of 0.857 which is above average. It can be deduced that the higher the distinctive difference between the mean average, the higher the level of agreement between the raters, and the lower the distinctive difference between the mean average, the lower the level of agreement between the raters. Generally, the performance of the models built with dataset DS2 is better compared to the other datasets because it contains homogeneous compounds.

The evaluation criteria, that is, accuracy, sensitivity, specificity, precision, recall, and, f-measure, were used as raters in ranking the ensembles and, considering the high level of agreement obtained between the raters (the higher Kendall's W is, the higher the level of agreement), it shows the effectiveness of the ranking method.

The performance of the ensembles in the prediction of bioactive molecules, and the ranking of the ensembles using Kendall's coefficient of concordance shows that Stacked ensemble has better performance compared to other ensembles in the prediction of heterogenous molecules. Vote ensemble makes predictions based on the output of the base classifiers, but stacked passes the output from the base classifier to a meta classifier. This meta classifier tries to correct the wrong prediction made by the base classifiers. This in turn strengthens and improves the overall prediction performance.

V. CONCLUSION

Although stacked ensemble performed better than other ensembles against which it was compared with an accuracy of 9.7002% and 94.9007% for datasets DS1 and DS3 respectively, the performance was lower than the performance of voting and adaboost, with an accuracy of 98.2260% using dataset DS2, which contained structurally homogeneous bioactive molecules. It is therefore recommended that implementation should be tried with other chemical datasets, and Stacked ensemble be used for prediction of structurally heterogeneous bioactive molecules. The three different datasets, which are both heterogeneous and homogeneous, implemented in this study were all retrieved from the MDL Drug Data Report (MDDR) database. It will be worthwhile to implement the ensemble methods on datasets from other repositories. Also, further study to improve the performance of base classifiers, which in turn improves the performance of the ensembles used in predicting bioactive molecules is encouraged, since the combination of the classifiers produce the ensemble, which is expected to perform better than the average classifiers [47], [48].

REFERENCES

[1] J. Azmir, I. S. M. Zaidul, M. M. Rahman, K. M. Sharif, A. Mohamed, F. Sahena, M. H. A. Jahurul, K. Ghafoor, N. A. N. Norulaini, and A. K. M. Omar, "Techniques for extraction of bioactive compounds from plant materials: A review," *J. Food Eng.*, vol. 117, no. 4, pp. 426–436, 2013.

[2] S. S. Elhady, A. M. El-Halawany, A. M. Alahdal, H. A. Hassanean, and S. A. Ahmed, "A new bioactive metabolite isolated from the red sea marine sponge hyrtios erectus," *Molecules*, vol. 21, no. 1, p. 82, 2016.

[3] N. H. Shady, E. M. El-Hossary, M. A. Fouad, T. A. M. Gulder, M. S. Kamel, and U. R. Abdelmohsen, "Bioactive natural products of marine sponges from the genus hyrtios," *Molecules*, vol. 22, no. 5, p. 781, 2017.

[4] K. El-Mostafa, Y. El Kharrassi, A. Badreddine, P. Andreoletti, J. Vamecq, M. El Kebbaj, N. Latruffe, G. Lizard, B. Nasser, and M. Cherkaoui-Malki, "Nopal cactus (Opuntia ficus-indica) as a source of bioactive compounds for nutrition, health and disease," *Molecules*, vol. 19, no. 9, pp. 14879–14901, 2014.

[5] A. Rahman and B. Verma, "Ensemble classifier generation using non-uniform layered clustering and genetic algorithm," *Knowl.-Based Syst.*, vol. 43, no. 2, pp. 30–42, May 2013.

[6] L. T. Afolabi, F. Saeed, H. Hashim, and O. O. Petinrin, "Ensemble learning method for the prediction of new bioactive molecules," *PLoS ONE*, vol. 13, no. 1, 2018, Art. no. e0189538.

[7] H. Chen, J. Poon, S. K. Poon, L. Cui, K. Fan, and D. M.-Y. Sze, "Ensemble learning for prediction of the bioactivity capacity of herbal medicines from chromatographic fingerprints," *BMC Bioinf.*, vol. 16, no. 12, p. S4, 2015.

[8] D. H. Wolpert, "Stacked generalization," *Neural Netw.*, vol. 5, no. 2, pp. 241–259, 1992.

[9] L. He, Z. Yang, Z. Zhao, H. Lin, and Y. Li, "Extracting drug-drug interaction from the biomedical literature using a stacked generalization-based approach," *PLoS ONE*, vol. 8, no. 6, 2013, Art. no. e65814.

[10] F. Anifowose, J. Labadin, and A. Abdulraheem, "Improving the prediction of petroleum reservoir characterization with a stacked generalization ensemble model of support vector machines," *Appl. Soft Comput. J.*, vol. 26, pp. 483–496, Jan. 2015.

[11] A. Gupta and A. R. Thakkar, "Optimization of stacking ensemble configuration based on various metahueristic algorithms," in *Proc. IEEE Int. Adv. Comput. Conf. (IACC)*, Feb. 2014, pp. 444–451.

[12] P. Shunmugapriya and S. Kanmani, "Optimization of stacking ensemble configurations through artificial bee colony algorithm," *Swarm Evol. Comput.*, vol. 12, no. 12, pp. 24–32, Oct. 2013.

[13] K. Dinakar, E. Weinstein, H. Lieberman, and R. L. Selman, "Stacked Generalization Learning to Analyze Teenage Distress," in *Proc. 8th Int. AAAL Conf. Weblogs Social Media*, 2014, pp. 81–90.

[14] L. F. Nicolas-Alonso, R. Corralejo, J. Gomez-Pilar, D. Álvarez, and R. Hornero, "Adaptive stacked generalization for multiclass motor imagery-based brain computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 23, no. 4, pp. 702–712, Jul. 2015.

[15] D. P. Gaikwad and R. C. Thool, "Intrusion detection system using bagging with partial decision treebase classifier," *Procedia Comput. Sci.*, vol. 49, pp. 92–98, Jan. 2015.

[16] A. Mert, N. Kılıç, and A. Akan, "Evaluation of bagging ensemble method with time-domain feature extraction for diagnosing of arrhythmia beats," *Neural Comput. Appl.*, vol. 24, no. 2, pp. 317–326, 2014.

[17] F. Moretti, S. Pizzuti, S. Panzieri, and M. Annunziato, "Urban traffic flow forecasting through statistical and neural network bagging ensemble hybrid modeling," *Neurocomputing*, vol. 167, pp. 3–7, Nov. 2015.

[18] H. M. Ashtawy and N. R. Mahapatra, "BgN-score and BsN-score: Bagging and boosting based ensemble neural networks scoring functions for accurate binding affinity prediction of protein-ligand complexes," *BMC Bioinf.*, vol. 16, p. S8, Feb. 2015.

[19] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *Proc. Int. Conf. Mach. Learn.*, 1996, pp. 148–156.

[20] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, Aug. 1997.

[21] B. Niu, Y. Jin, W. Lu, and G. Li, "Predicting toxic action mechanisms of phenols using AdaBoost learner," *Chemometrics Intell. Lab. Syst.*, vol. 96, no. 1, pp. 43–48, 2009.

[22] S. K. Mathanker, P. R. Weckler, T. J. Bowser, N. Wang, and N. O. Maness, "AdaBoost classifiers for pecan defect classification," *Comput. Electron. Agricult.*, vol. 77, no. 1, pp. 60–68, 2011.

[23] Y.-T. Yan, Y.-P. Zhang, J. Chen, and Y.-W. Zhang, "Incomplete data classification with voting based extreme learning machine," *Neurocomputing*, vol. 193, pp. 167–175, 2016.

[24] B. Al-Shboul, H. Hakh, H. Faris, I. Aljarah, and H. Alsawalqah, "Voting-based classification for e-mail spam detection," *J. ICT Res. Appl.*, vol. 10, no. 1, pp. 29–42, 2016.

[25] A. Husain and M. H. Khan, "Early diabetes prediction using voting based ensemble learning," *Adv. Comput. Data Sci.*, vol. 905, pp. 95–103, Oct. 2018.

[26] O. O. Petinrin and F. Saeed, "Bioactive molecule prediction using majority voting-based ensemble method," *J. Intell. Fuzzy Syst.*, vol. 35, no. 1, pp. 383–392, 2018.

[27] B. Gopinath and B. R. Gupt, "Majority voting based classification of thyroid carcinoma," *Procedia Comput. Sci.*, vol. 2, pp. 265–271, Jan. 2010.

[28] I. Gandhi and M. Pandey, "Hybrid Ensemble of classifiers using voting," in *Proc. Int. Conf. Green Comput. Internet Things (ICGCIoT)*, Oct. 2015, pp. 399–404.

[29] M. S. Zia, M. Hussain, and M. A. Jaffar, "A novel spontaneous facial expression recognition using dynamically weighted majority voting based ensemble classifier," *Multimedia Tools Appl.*, vol. 77, no. 19, pp. 25537–25567, Oct. 2018.

[30] A. Onan, S. Korukoğlu, and H. Bulut, "A multiobjective weighted voting ensemble classifier based on differential evolution algorithm for text sentiment classification," *Expert Syst. Appl.*, vol. 62, pp. 1–16, Nov. 2016.

[31] R. Polikar, *Ensemble Learning* (Ensemble Machine Learning). Boston, MA, USA: Springer, 2012, pp. 1–34.

[32] F. Saeed, N. Salim, and A. Abdo, "Voting-based consensus clustering for combining multiple clusterings of chemical structures," *J. Cheminform.*, vol. 4, no. 1, 2012, Art. no. 37.

[33] F. Saeed, A. Ahmed, M. S. Shamsir, and N. Salim, "Weighted voting-based consensus clustering for chemical structure databases," *J. Comput.-Aided Mol. Des.*, vol. 28, no. 6, pp. 675–684, 2014.

[34] D. Guan, W. Yuan, T. Ma, and S. Lee, "Detecting potential labeling errors for bioinformatics by multiple voting," *Knowl.-Based Syst.*, vol. 66, pp. 28–35, Aug. 2014.

[35] J. H. Hu, Y. Li, J.-Y. Yang, H.-B. Shen, and D.-Y. Yu, "GPCR–drug interactions prediction using random forest with drug-association-matrix-based post-processing procedure," *Comput. Biol. Chem.*, vol. 60, pp. 59–71, Feb. 2016.

[36] Z. Mašetic and A. Subasi, "Congestive heart failure detection using random forest classifier," *Comput. Methods Programs Biomed.*, vol. 130, pp. 54–64, Jul. 2016.

[37] T. G. Pavey, N. D. Gilson, S. R. Gomersall, B. Clark, and S. G. Trost, "Field evaluation of a random forest activity classifier for wrist-worn accelerometer data," *J. Sci. Med. Sport*, vol. 20, no. 1, pp. 75–80, 2017.

[38] T. Shaikhina, D. Lowe, S. Daga, D. Briggs, R. Higgins, and N. Khovanova, "Decision tree and random forest models for outcome prediction in antibody incompatible kidney transplantation," *Biomed. Signal Process. Control*, vol. 52, pp. 456–462, Jul. 2019.

[39] G. Cano, J. Garcia-Rodriguez, A. Garcia-Garcia, H. Perez-Sanchez, J. A. Benediktsson, A. Thapa, and A. Barr, "Automatic selection of molecular descriptors using random forest: Application to drug discovery," *Expert Syst. Appl.*, vol. 72, pp. 151–159, Apr. 2017.

[40] J.-H. Huang, M. Wen, L.-J. Tang, H.-L. Xie, L. Fu, Y.-Z. Liang, and H.-M. Lu, "Using random forest to classify linear B-cell epitopes based on amino acid properties and molecular features," *Biochimie*, vol. 103, pp. 1–6, Aug. 2014.

[41] A. V. Lebedev, E. Westman, G. J. P. Van Westen, M. G. Kramberger, A. Lundervold, D. Aarsland, H. Soininen, I. Kłoszewska, P. Mecocci, M. Tsolaki, B. Vellas, S. Lovestone, and A. Simmons, "Random forest ensembles for detection and prediction of Alzheimer's disease with a good between-cohort robustness," *NeuroImage Clin.*, vol. 6, pp. 115–125, Jan. 2014.

[42] A. Abdo and N. Salim, "Ligand-based virtual screening using Bayesian inference network," *Library Des. Search Methods Appl.*, vol. 8, no. 4, pp. 57–59, 2011.

[43] A. Abdo, F. Saeed, H. Hamza, A. Ahmed, and N. Salim, "Ligand expansion in ligand-based virtual screening using relevance feedback," *J. Comput. Aided. Mol. Des.*, vol. 26, no. 3, pp. 279–287, 2012.

[44] M. M. Al-Dabbagh, N. Salim, M. Himmat, A. Ahmed, and F. Saeed, "A quantum-based similarity method in virtual screening," *Molecules*, vol. 20, no. 10, pp. 18107–18127, 2015.

[45] A. Abdo, V. Leclère, P. Jacques, N. Salim, and M. Pupin, "Prediction of new bioactive molecules using a Bayesian belief network," *J. Chem. Inf. Model.*, vol. 54, no. 1, pp. 30–36, 2014.

[46] I. B. Mustapha and F. Saeed, "Bioactive molecule prediction using extreme gradient boosting," *Molecules*, vol. 21, no. 8, p. 983, Jul. 2016.

[47] F. Anifowose, J. Labadin, and A. Abdulraheem, "Ensemble model of artificial neural networks with randomized number of hidden neurons," in *Proc. 8th Int. Conf. Inf. Technol. Asia (CITA)*, Jul. 2013, pp. 1–5.

[48] S. Džeroski and B. Ženko, "Is combining classifiers with stacking better than selecting the best one?" *Mach. Learn.*, vol. 54, no. 3, pp. 255–273, 2004.

OLUTOMILAYO OLAYEMI PETINRIN received the B.Sc. degree in computer science from Ekiti State University, Ado Ekiti, Nigeria, and the M.Sc. degree in computer science from Universiti Teknologi Malaysia. She is currently pursuing the Ph.D. degree with the City University of Hong Kong. She has published few articles in the area of cheoinformatics and bioinformatics. Her research interests include data mining, machine learning, artificial intelligence, bioinformatics, and chemoinformatics.

FAISAL SAEED received the B.Sc. degree in computers (information technology) from Cairo University, Egypt, and the M.Sc. degree in information technology management and the Ph.D. degree in computer science from Universiti Teknologi Malaysia (UTM), Malaysia. He was a Senior Lecturer with the Department of Information Systems, Faculty of Computing, UTM. He has been an Assistant Professor with the Information Systems Department, Taibah University, KSA, since 2017. He published more than 80 articles in indexed journals and international conferences. His research interests include data mining, information retrieval, and machine learning. He served as a General Chair of several international conferences and as a Guest Editor in more than 15 Scopus and ISI Journals.

● ● ●