Journal of Modern Computing and Engineering Research
JMCER
Volume 2022, p. 94-100
https://jmcer.org

Performance Evaluation of the Ensemble and Selected Machine Learning Techniques

Nadia Mohammed Majeed[1] and Fawziya Mahmood Ramo[2]

[1,2]*Department of Computer Science, College of Computer Science and Mathematics, University of Mosul, 41002 Mosul, Iraq*
[1]*nadia.20csp72@student.uomosul.edu.iq*, [2]*fawziyaramo@uomosul.edu.iq*

Received: November 25, 2022 *Revised: December 27, 2022* *Accepted: December 28, 2022*

Abstract

Ensemble methods are machine-learning techniques that include the creation of several learners for a given task. Ensemble techniques aim to achieve high classification accuracy and improve performance. In predicting breast cancer, we require enhancing the accuracy of algorithms; therefore, we utilize here an ensemble technique that combines predictions of several models. In this study, the proposed ensemble hard voting classifier employs a combination of five machine learning algorithms: Support Vector Machine (SVM), K-Nearest Neighbours(K-NN), Naive Bayes (NB), Decision Tree (DT), and Random Forest (RF) is used to provide a binary classification for breast cancer. The results of the individual classifiers are then combined and compared with the performance of five individual classifiers with the hard voting classifier. The results show that ensemble-voting techniques perform better than single classifiers. The Wisconsin Breast Cancer Dataset (WBCD) from the UCI machine-learning repository was used in our experiments. The proposed ensemble hard voting classifier has given the highest accuracy value with 96.49%, whereas Support Vector Machine, Nearest Neighbours, Naive Bayes, Decision Tree, and Random Forest achieved accuracies of 95.32%, 92.39%, 94.73%, 92.98%, and 95.32% respectively on the breast cancer dataset.

Keywords: *Ensemble Learning, Hard Voting, machine learning, WBCD dataset (Wisconsin Breast Cancer Dataset).*

1 Introduction

An ensemble technique is a combination of machine learning methods, the outputs of which are aggregated to produce the classification process outcome. Machine learning used for decades to diagnose cancer, but as the disease's stages and complexity rise, clinicians require new developments in the field to make reliable predictions about the disease's course. The proposed work aims to predict breast cancer using an ensemble model and machine learning approaches. SVM, DT, NB, K-NN, Perceptron, Logistic Regression, and other algorithms previously used to predict breast cancer. When compared to ensemble methods, these individual classifiers have low performance (Leena Nesamani and Nirmala Sugirtha Rajini, 2020) Five supervised machine-learning algorithms are included in our model are SVM, K-NN, NB, DT, and RF, for better performance and reliable classification accuracy all of these classifiers have been applied to the ensemble hard voting method. On the publicly accessible WBCD. Accuracy, F-score, and other metrics were used to assess the performances of the ensemble and solo models (Saad Assiri, Nazir, and Velastin, 2019).

2 Related Works

In recent years, many studies conducted on the subject of ensemble learning, and multiple types of research presented in this regard. One of the most important of these studies:

Rathore et al. (Rathore, Divya, and Agarwal, 2014) used an ensemble technique For forecasting breast cancer. The ensemble classifier based on the voting method is built using the (DT) Classifier, (NB), and Classification is based on Multiple Association Rules (CMAR) classifier. In comparison to the traditional classifier, experimental findings show that the suggested technique achieves higher accuracy. Breast cancer data were obtained from SEER (Surveillance of Epidemiology and End Result) in which the National Cancer Institute participates.

Kumar et al. (Kumar, Nikhil, and Sumangali, 2017) used the three best supervised educational

Corresponding author: nadia.20csp72@student.uomosul.edu.iq

classification techniques for breast cancer detection such as J48, Naïve Bayes, and SVM were used and compared on different criteria. The study showed that aggregating all three methods utilizing an ensemble voting classifier is better for breast cancer detection. The data set is from the University of Wisconsin database.

Gupta, Madhuri and Bharat Gupta (Gupta and Gupta, 2018) proposed four machine-learning approaches utilizing the WBCD. For the analysis, the PCA identified a total of twenty features. To classify the data separately, the model employed machine learning methods like SVM, Logistic Regression, DT, and K–NN. The ensemble voting methods were utilized to aggregate the outcomes of the different methods. This system attempts to forecast the target class for every input data depending on the weights given to the algorithms by the Sequential Least Squares Programming Method (SLSQP). Their experiment shows that the voting classifier improves ensemble system performance when compared to individual classifiers, with a 97.88 %, whereas the SVM, KNN, DT, and Logistic Regression yield an accuracy of 93.98 %, 90.12 %, 92.15 %, and 89.12 %, respectively in the ensemble model.

Nguyenet et al. (Nguyen *et al.*, 2019) use the Wisconsin Breast Cancer Dataset to evaluate both supervised and unsupervised strategies for breast cancer detection. To predict breast cancer, classifiers such as K-NN, SVM, LR, AdaBoost, Perceptron, XGBoost, Extremely Randomized Trees, and Gradient Descent were utilized. The outcomes of the separated classifiers were integrated using Ensemble Voting algorithms. Equal weights are given to all classifiers to guarantee that they all have the same desire to take part in the voting method. Of all the methods used for prediction, only four performed the best, with an accuracy of about 98 %, they are Voting Classifier, LR, SVM, and, AdaBoost.

Raza, Khalid (Raza, 2019) suggested an ensemble model that aggregates the output of three classifiers: LR, multilayer perceptron, and NB, with the majority voting approach used to predict heart disease. The proposed ensemble approach outperformed any single classification technique with an accuracy of 88.88%.

MurtiRawat et al. (Murtirawat *et al.*, 2020) suggested various machine learning methods to assist in the detection of breast cancer, including (LR), (KNN), and voting classifier with Principal Component Analysis (PCA).WBCD was utilized to train and test the classifiers. Before utilizing PCA to extract features from the data set, the data was pre-processed. The outcome of the suggested approach shows a classification accuracy of 98.60 % when using K-NN and 97.90 % when using LR, whereas the voting classifier has the greatest accuracy of 99.30 %.

Assiri et al. (Adel S. Assiri, Nazir, and Velastin, 2020) basic LR, SVM with stochastic gradient descent, and multilayer perceptron network are three classifiers suggested for ensemble algorithms utilizing a voting classifier. For the hard vote classifier, a majority-based voting system was utilized. The hard voting method outperforms the individual technique with 99.42 % for the WBCD.

GurjotKour et al. (Abrol, Kalrupia, and Kaur, 2022) proposed Voting Classifier for Corona Infection detection The obtained dataset in this study comprises COVID-19 instances from Mexico. The dataset is additionally handled for feature reduction using the PCA technique, followed by the use of the k-means method, which can cluster similar and different characteristics. For COVID-19 detection, the voting algorithm is used, which is an aggregation of NB, RF, SVM, and Bernoulli naive Bayes. The findings indicate that Logistic Regression has an accuracy of 84 %, NB has an accuracy of 82 %, and the voting classification technique has the highest accuracy of 94 %.

3 Contribution

The major goal of this paper is to build an ensemble technique for accurately predicting breast cancer. The cancer dataset WBCD is used as input in the approach. With the use of the set union operation, feature selection methods are used to choose the most important attributes. Classification algorithms are the evaluation process for a specific dataset in sequential order. A split test train was used to compare the outcomes. Finally, the outcomes are voted by hard voting to determine the result.

4 Machine Learning Techniques

4.1 Support Vector Machine (SVM)

Is a technique for machine learning that categorizes data by identifying the best hyperplane. (Li and Chen, 2020a) Several continuous and categorical variables can be handled with ease. To distinguish between different classes, SVM generates a hyperplane in multidimensional space. To decrease error, SVM iteratively creates an optimum hyperplane (Navlani, 2019). To achieve maximum generalization, the SVM uses the training dataset's largest classification margin as the decision boundary and divides the data into two groups (Benign and Malignant). The model that results may be used to

predict the category of new samples(Li and Chen, 2020a).

- Support vectors:

The data points which are near the hyperplane are known as support vectors. By using margin calculation, these points will best identify the dividing line. These points are important to the creation of the classifier.

- Hyperplane:

Is a decision plane that distinguishes between a collection of objects with various category memberships.

- Margin:

Is the space between the two lines at the closest class points. It is determined by measuring the perpendicular distance between the line and the nearest points or support vectors. It is regarded to be a better margin if there is a bigger gap between the classes; otherwise, it is considered to be a bad margin. The following are the steps that SVM performs to discover the biggest marginal hyperplane:

A. Generate hyperplanes that divide classes. in the better way possible. Three hyperplanes are seen in the figure on the left, blue, black, and orange. In this, case the black is dividing the two classes properly while blue and orange have a larger categorization error
B. Choose the appropriate hyperplane, as shown in the figure on the right, with the greatest separation from any nearest data points.

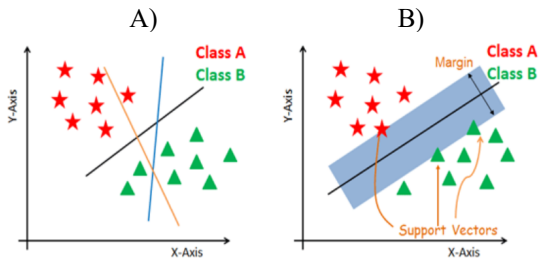

Figure 1: A) Create hyperplanes and B) Choose the appropriate hyperplane (Navlani, 2019)**.**

4.2 K-Nearest Neighbors

Is a simple approach that retains all of the available cases and a similarity metric is used to classify new ones. A case is categorized by a majority vote among its neighbors, and it is then classified as the most common class among those closest to it. This is measured using a distance function. For example, If K=1, the case is classified to the class of its closest neighbor(Iqbal, Nassif, and Shahin, 2020a). KNN is very much useful in the field of classification techniques, because of: strong to noisy training data and effectiveness if the training data is vast(Prince, Hasan, and Shah, 2019). Its simplicity and superior precision were employed in a variety of data processing applications, like data mining, pattern recognition, and machine learning. As one of the top ten data mining algorithms(Kumari and Philosophy, 2012; Thirumal and Nagarajan, 2015).

When a new data point, x1, is present and there are two categories, A and B, which one does it belong in? The K-NN algorithm is needed to handle this kind of issue K-NN makes determining the category of a given dataset straightforward. See the figure below:

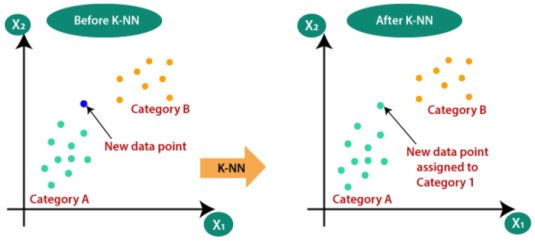

Figure 2: determine the category of a specific dataset

How does operate?

The function of the K-NN can be clarified based on the following stages:

Stage 1: Choose the K-number of the neighbors.

Stage 2: compute the K-neighbors' Euclidean distance

Stage 3: Choose the K closest neighbors depending on computed Euclidean distance.

Stage 4: Determine how many data items are in every category among k neighbors.

Stage 5: Put the new data in the category with the greatest number of neighbors

Stage 6: The entire model is prepared (javatpoint, 2021a)**.**

4.3 Naïve Bayes

Is a method for categorizing data, well known for ease of use and efficiency. It is also fast to construct and predicts quickly. The Naive Bayes method is a probabilistic categorization algorithm that learns feature probabilities depending on the target class. It is assumed that the existence of one feature does not affect the existence of the others Even if the other feature is important, Nave Bayes can perform better because it does not require precise probabilities estimations as long as the best probability is assigned to the proper class. It is built on the Bayes theorem, which explains:

$$P(c \mid x) = \frac{P(x \mid c)\, P(c)}{p(x)} \qquad (1)$$

Where $P(c \mid x))$ *and* $P(x \mid c)$ are the conditional probability of an event occurring. assuming that event x is correct and vice versa. (Jafari and Olbe, 2021) .

These stages will provide you with the groundwork necessary to execute a Naive Bayes implementation and use it to solve your forecasting model issues.

Stage 1: Split by Class.
Stage 2: Summarize the dataset.
Stage 3: Data Summaries by Category.
Stage 4: The probability density function of Gaussian.
Stage 5: likelihood by class. ('Naive Bayes Classifier From Scratch in Python - by J.Brownlee - 2019', no date)

4.4 Decision Tree (DT)

Is also a popular Classification technique because it is similar to human thinking and simple to learn. The idea behind the DT is to utilize a tree diagram to explain the method of several decision tree rules from a dataset with features and labels. A new sample may be easily categorized using an existing decision tree, using the fundamental principle of constantly matching the corresponding features and related conditions until reaching a leaf node (Figure.3), in this way, the class label's leaf node can be utilized as the sample. The decision tree's computing complexity is low, especially when evaluating new data (Li and Chen, 2020b).

The following stages will help you comprehend the entire process:

Stage 1: begin the tree at the root node S, which includes the whole dataset.
Stage 2: Utilize Attribute Selection Measure (ASM) to determine the dataset's better attribute.
Stage 3: Separate the S into categories that include potential values for better.
Stage 4: Create the decision tree node that includes the best features
Stage 5: The subsets of the dataset generated in step 3 will be used to iteratively form a new decision tree. Keep going through this manner until you are unable to do so any longer categorize the nodes, at which point the last node will be referred to as a leaf node (javatpoint, 2021b).

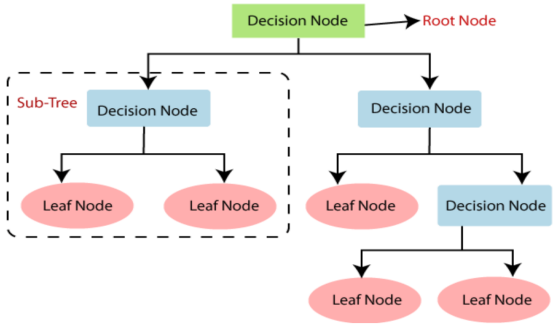

Figure 3: The overall form of a decision tree (javatpoint, 2021b)

4.5 Random Forest (RF)

Is an ensemble machine learning method. (Adel S Assiri, Nazir and Velastin, 2020). Is composed of a large number of multiple decision trees. Each tree gives a class prediction, and the model's prediction is determined by the class with the most votes(Iqbal, Nassif, and Shahin, 2020b). This approach employs a "parallel ensemble," which involves fitting multiple decision tree algorithms in parallel (Sarker, 2021). When a large volume of data needs be to needs categorized based on accuracy and correctness, decision trees are an excellent alternative. To obtain accuracy in the prediction you must offer information about the input variables. The use of the random forest is strongly suggested for categorizing data with good precision, although some data in the dataset is missing, and maintaining accuracy when the amount of data grows fast. In many cases, random forest techniques seem to be more accurate than single classifiers (Adel S Assiri, Nazir, and Velastin, 2020).

The random forest algorithm's stages are described as follows:

Stage 1: With RF, n records are randomly chosen from a data set involving k records.
Stage 2: An individual decision tree is generated for each sample.
Stage 3: Each decision tree will provide an output.
Stage 4: The outcome for classification is based on either majority or voting averaging (Sruthi, 2021).

5 Ensemble Classification

Ensemble learning combines multiple machine learning models to improve performance (Ibrahim, Nazir, and Velastin, 2021). Sometimes the amount of data on which to perform learning is so large that a single classifier cannot handle or assess it. This large amount of data may be examined by dividing it up into smaller parts and distributing it to multiple

classifiers (ensemble) instead of just one field (AL-Malali, 2021).

Ensemble learning is separated into three kinds:

1. Voting ensemble learning.
2. Bagging ensemble learning
3. Boosting ensemble learning

One of the easiest ensemble learning strategies is the voting method, which combines predictions from many models with either a hard or a soft vote. Hard voting ensemble learning was applied in our research which combined the result of prediction for five individual classifiers as illustrated in (figure.4). Hard voting is the easiest kind of majority voting in which the largest number of votes is expected(Ibrahim, Nazir, and Velastin, 2021). Every single classifier provides its prediction. The model then selects the classifier with the most significant number of prediction outcomes (Fitni and Ramli, 2020). Being the most often predicted target class by the classifiers. The target class y is predicted using majority voting by the classifiers C (Ibrahim, Nazir and Velastin, 2021):

$$y = mode\ \{C1(x).C2(x).\ldots\ Cn(x)\} \qquad (2)$$

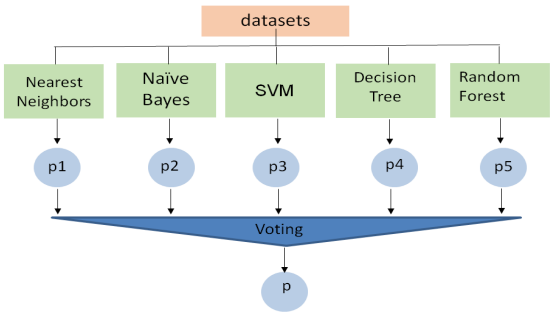

Figure 4: Overall Structure ensemble voting classifier (Das and D. Biswas, 2019)

6 Result and Discussion

The suggested work was carried out using the Python programming language. A novel ensemble classifier that combines individual classifiers like, SVM, K-NN, NB, DT, and RF all of these classifiers has been applied to the ensemble Voting Classifiers utilized in breast cancer diagnosis using the WBCD dataset. This dataset contains 569 instances and each instance has 30 attributes taken from nuclei images, and it is divided into testing and training. In the training phase, the relevant data is entered into the Algorithms, amounting to (70%) of the total number of the database, amounting to (398). This data is divided into two categories, the first category contains images (benign tumors), and the second category contains images (malignant tumors). While the testing phase is intended to ensure the performance of the proposed models. After completing the training process for the algorithms and their stability, they were tested on a set of test data amounting to (30%) of the total database, which is (171) images, which are also divided into two categories, benign and malignant. Individual classifiers that are combined using Majority Voting will produce better results than any other individual models used in the evaluation which is equal to (96.49122) from prediction using single algorithms (SVM, K-NN, NB, DT, and RF) whose accuracy is 95.32163, 92.39766, 94.73684, 92.98245, 95.32163, respectively. As shown in the table below.

Table 1: Individual classifiers obtained from Prediction metrics.

Models	Accuracy	Precision	Recall	F1-Score
SVM	95.32163	94.64285	98.14814	96.36363
K-NN	92.39766	92.79279	95.37037	94.06392
Naïve Bayes	94.73684	95.41284	96.29629	95.85253
Decision Tree	92.98245	90.67796	99.07407	94.69026
Random Forest	95.32163	95.45454	97.22222	96.33027
Ensemble Voting Classifier	96.49122	95.53571	99.07407	97.27272

A set of metrics was employed to assess the performance of our model; the prediction metrics utilized are Accuracy, Precision, Recall, and F-measure These metrics are imported from the "sklearn" package in python. The following are their definitions:

Accuracy is an excellent fundamental statistic to evaluate the effectiveness of the model, used to determine the number of correct predictions the model made throughout the whole test dataset.

The precision determines how many correctly forecasted cases turned out to be positive. This would decide the reliability of our model.

Recall demonstrates how many actual positive instances we were able to correctly predict using our model. When False Negative outperforms False Positive, the recall metric can be helpful.

The recall decreases when we attempt to enhance model precision and vice versa. A combination idea on these two measures is provided by the F1-Score, which is a harmonic mean of precision and recall. It is maximum, when recall and precision are equal (Karimi, 2021).

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

$$F1 - Score = \frac{(2 * Precision * Recall)}{(Precision + Recall)} \quad (6)$$

Where TP: True Positive, FP: False Positive, TN: True Negative, and FN: False Negative (AL-Malali, 2021).

All of the classification models' positive class was discovered to be 'benign.' The number of positive or benign breast cancer cases is referred to as TP, whereas the number of negative or malignant instances is referred to as FP. TN denotes instances that were accurately identified as not benign (malignant), whereas FN denotes cases that were incorrectly categorized as malignant.

7 Conclusion

To predict breast cancer, several different classifiers were used. And the top classifiers with the highest prediction accuracy were chosen to produce an ensemble model, in which the individual model's findings were combined to provide the final prediction utilizing majority voting techniques. The ensemble models' final forecasts were better than the individual classifier's predictions, indicating that ensemble techniques are good predictors of breast cancer and may be used for any categorization task. Ensemble voting improves the system's stability and performance. As a result, the suggested technique is more robust to any unexpected instances that may occur in any classifier model. As a result, our suggested ensemble-based method is more feasible, and it would provide superior treatment and precise diagnosis for patients with breast cancer. In future work, we want to implement our module using deep learning techniques.

References

Abrol, P., Kalrupia, N. and Kaur, J. (2022) 'Hybrid Voting Classifier Model for COVID-19 Prediction by Embedding Machine Learning Techniques', 13(02), pp. 171–183.

AL-Malali, M. K. H. (2021) 'Behavioral Sense Classification using Machine Learning Algorithms', pp. 1–144.

Das and D. Biswas (2019) 'Prediction of breast cancer using ensemble learning', in *2019 5th International Conference on Advances in Electrical Engineering, ICAEE 2019*. Institute of Electrical and Electronics Engineers Inc., pp. 804–808. doi: 10.1109/ICAEE48663.2019.8975544.

Assiri, Adel S, Nazir, S. and Velastin, S. A. (2020) 'Breast tumor classification using an ensemble machine learning method', *Journal of Imaging*, 6(6), p. 39.

Assiri, Adel S., Nazir, S. and Velastin, S. A. (2020) 'Breast Tumor Classification Using an Ensemble Machine Learning Method', *Journal of Imaging*, 6(6). doi: 10.3390/JIMAGING6060039.

Fitni, Q. R. S. and Ramli, K. (2020) 'Implementation of ensemble learning and feature selection for performance improvements in anomaly-based intrusion detection systems', *Proceedings - 2020 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology, IAICT 2020*, pp. 118–124. doi: 10.1109/IAICT50021.2020.9172014.

Gupta, M. and Gupta, B. (2018) 'An Ensemble Model for Breast Cancer Prediction Using Sequential Least Squares Programming Method (SLSQP)', *2018 11th International Conference on Contemporary Computing, IC3 2018*, pp. 1–3. doi: 10.1109/IC3.2018.8530572.

Ibrahim, S., Nazir, S. and Velastin, S. A. (2021) 'Feature selection using correlation analysis and principal component analysis for accurate breast cancer diagnosis', *Journal of Imaging*, 7(11). doi: 10.3390/jimaging7110225.

Iqbal, H. N., Nassif, A. B. and Shahin, I. (2020a) 'Classifications of Breast Cancer Diagnosis using Machine Learning', *International Journal of Computers*, 14(January 2021), pp. 86–86. doi: 10.46300/9108.2020.14.13.

Iqbal, H. N., Nassif, A. B. and Shahin, I. (2020b) 'Classifications of Breast Cancer Diagnosis using Machine Learning', *International Journal of Computers*, 14(January 2021), p. 86. doi: 10.46300/9108.2020.14.13.

Jafari, M. and Olbe, J. (2021) 'A Comparison of Machine Learning Algorithms for Predicting Winners of League of Legends Matches'.

javatpoint (2021a) 'K-Nearest Neighbor(KNN) Algorithm for Machine Learning - Javatpoint', *Javatpoint*. Available at: https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning.

javatpoint (2021b) 'Machine Learning Decision Tree Classification Algorithm - Javatpoint', *Javatpoint.Com*. Available at: https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm.

Karimi, Z. (2021) 'Confusion Matrix', (October).

Kumar, U. K., Nikhil, M. B. S. and Sumangali, K. (2017) 'Prediction of breast cancer using voting classifier technique', *2017 IEEE International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials, ICSTM 2017 - Proceedings*, (August), pp. 108–114. doi: 10.1109/ICSTM.2017.8089135.

Kumari, P. K. and Philosophy, D. O. F. (2012) 'a Study on Novel Techniques for Heart Sound and Murmur Classification , Search and Retrievel', pp. 1–174.

Leena Nesamani, S. and Nirmala Sugirtha Rajini, S. (2020) 'Evaluation of ensemble machines in breast cancer prediction', *Advances in Parallel Computing*, 37(1), pp. 391–395. doi: 10.3233/APC200173.

Li, Y. and Chen, W. (2020a) 'A comparative performance assessment of ensemble learning for credit scoring', *Mathematics*, 8(10), pp. 1–19. doi: 10.3390/math8101756.

Li, Y. and Chen, W. (2020b) 'A comparative performance assessment of ensemble learning for credit scoring', *Mathematics*, 8(10), p. 1756.

Murtirawat, R. *et al.* (2020) 'Breast Cancer Detection Using K-Nearest Neighbors, Logistic Regression and Ensemble Learning', *Proceedings of the International Conference on Electronics and Sustainable Communication Systems, ICESC 2020*, (Icesc), pp. 534–540. doi: 10.1109/ICESC48915.2020.9155783.

J.Brownlee - 2019' 'Naive Bayes Classifier From Scratch in Python (no date). Available at: https://machinelearningmastery.com/naive-bayes-classifier-scratch-python/.

Navlani, A. (2019) 'Support Vector Machines with Scikit-learn', *DataCamp*, pp. 1–15. Available at: https://www.datacamp.com/community/tutorials/svm-classification-scikit-learn-python.

Nguyen, Q. H. *et al.* (2019) 'Breast Cancer Prediction using Feature Selection and Ensemble Voting', *Proceedings of 2019 International Conference on System Science and Engineering, ICSSE 2019*, pp. 250–254. doi: 10.1109/ICSSE.2019.8823106.

Prince, M. S. M., Hasan, A. and Shah, F. M. (2019) 'An Efficient Ensemble Method for Cancer Detection', *1st International Conference on Advances in Science, Engineering and Robotics Technology 2019, ICASERT 2019*, 2019(Icasert), pp. 1–6. doi: 10.1109/ICASERT.2019.8934817.

Rathore, N., Divya and Agarwal, S. (2014) 'Predicting the survivability of breast cancer patients using ensemble approach', *Proceedings of the 2014 International Conference on Issues and Challenges in Intelligent Computing Techniques, ICICT 2014*, pp. 459–464. doi: 10.1109/ICICICT.2014.6781326.

Raza, K. (2019) *Improving the prediction accuracy of heart disease with ensemble learning and majority voting rule*, *U-Healthcare Monitoring Systems*. Elsevier Inc. doi: 10.1016/b978-0-12-815370-3.00008-6.

Saad Assiri, A., Nazir, S. and Velastin, S. A. (2019) 'A Hybrid Ensemble Method for Accurate Breast Cancer Tumor Classification using State-of-the-Art Classification Learning Algorithms', (November). Available at: www.preprints.org.

Sarker, I. H. (2021) 'Machine Learning: Algorithms, Real-World Applications and Research Directions', *SN Computer Science*, 2(3). doi: 10.1007/s42979-021-00592-x.

Sruthi, E. R. (2021) 'Random Forest | Introduction to Random Forest Algorithm', *AnalyticsVidya.com*. Available at: https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/.

Thirumal, P. C. and Nagarajan, N. (2015) 'Utilization of data mining techniques for diagnosis of diabetes mellitus - A case study', *ARPN Journal of Engineering and Applied Sciences*, 10(1), pp. 8–13.

Biography

Nadia Mohammed Majeed obtained her B.Sc. degree from Al-Hadbaa University College/ Mosul/ Iraq. Currently, she is an M.Sc. student in Computer Science at the University of Mosul / Iraq. She is interested in the AI field of research.

Fawziya Mahmood Ramo obtained her B.Sc. degree in Computer Science in 1992 and obtained her M.Sc. degree in Computer Science in 2001 and her Ph.D. in Artificial Intelligence in 2007 from the University of Mosul, Iraq. She obtained an academic position as an Assistant professor in 2013, then obtained a professor position in 2021. She is interested in research in computer science and artificial intelligence and machine learning.