IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 15, No. 3, June 2026, pp. 2410~2422
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i3.pp2410-2422

Ensuring effective cervical cancer diseases diagnosis system using ensemble machine learning model

Oluwatobi Akinlade[1], Jafar Abdollahi[2], Misan Paul Etchie[3], Sunday Adeola Ajagbe[4,5], Oluwaseyi Omotayo Alabi[6], Bambo Ayo Adeyanju[7]
[1]Department of Computer Science, Birmingham City University, Birmingham, United Kingdom
[2]Department of Computer Engineering, Faculty of Engineering, Central Tehran Branch, Islamic Azad University, Tehran, Iran
[3]School of Informatics, Computing and Cyber Systems, Northern Arizona University, Flagstaff, United States
[4]Department of Computer Science, University of Zululand, Richards Bay, South Africa
[5]Department of Computer Engineering, Abiola Ajimobi Technical University, Ibadan, Nigeria
[6]Department of Mechanical Engineering, Lead City University, Ibadan, Nigeria
[7]International Business School, University of Lincoln, Lincoln, United Kingdom

Article Info

Article history:

Received Aug 22, 2024
Revised Mar 3, 2026
Accepted Apr 22, 2026

Keywords:

Cervical cancer
Deep learning
Diagnosis
Ensemble models
Machine learning

ABSTRACT

Worldwide, cervical cancer is a major public health concern, and early detection is essential for successful management and treatment. Ensemble machine learning (ML) has emerged as a promising approach for improving the accuracy and reliability of cervical cancer diagnosis systems. This study evaluated the performance of various ensemble ML models for cervical cancer diagnosis using a large dataset of cervical cell images. The performance of different ensemble models was compared, including random forest (RF), gradient boosting (GB), and stacking, with conventional ML models, such as logistic regression and support vector machine (SVM). The bagging ensemble model developed reported training of 0.991667, 0.827586, 1.0, and 0.90566 for accuracy, precision, recall, and F-measure (F1-score), respectively. Furthermore, the interpretability of ensemble models was investigated using feature importance and partial dependence plots. The interpretability analysis revealed that ensemble models can provide valuable insights into the key features and factors that contribute to cervical cancer diagnosis. In conclusion, the findings suggest that ensemble ML is a promising approach for developing accurate and reliable cervical cancer diagnosis systems and improving the interpretability of these models.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Oluwaseyi Omotayo Alabi
Department of Mechanical Engineering, Lead City University
Toll-Gate Area, Off Oba Otudeko Avenue, Ibadan, Nigeria
Email: alabi.oluwaseyi@lcu.edu.ng

1. INTRODUCTION

Many studies have employed artificial intelligence (AI) techniques as tools to solve healthcare-related issues, particularly, and the performance has been so encouraging. Among these studies was heart disease prediction using internet of things (IoT)-enabled machine learning (ML) [1]. The automated diagnosis of infectious diseases using the IoT-enabled AI model has offered sustainable prognosis and reduced diagnosis errors significantly with improved accuracy [2]. The integration of IoT-enabled AI applications improves healthcare delivery in various ways. Although the classical deep learning (DL) techniques of AI in existence focus on automatic disease diagnosis systems where parameter setting, such as learning rate selection and weight management, are major concerns [3]–[5]. Meanwhile, limited attention is paid to the aspect of management using automated devices that integrated ensembled DL-based systems and IoT computing models for real-time disease diagnosis. The technologies' inherent advantages in the integration of an ensembled DL-based system and IoT computing have much power to diagnose many diseases at a time and aid in the management of disease diagnosis of diseases. Therefore, this research aims to evaluate ensemble ML-based systems for cervical cancer disease diagnosis. Specific objectives are to:

- To simulate the risk factors of cervical cancer diseases using ensemble ML-based models.
- To evaluate the ensemble ML-based models using various state-of-the-art metrics like accuracy, training time, sensitivity, specificity, recall, F-measure (F1-score), and precision.

This paper is structured as follows. Section 2 delineates the literature review pertinent to this study. Section 3 delineates the methodology, schematic framework, and data preparation of the sample dataset, and elucidates the diagnosis model, parameters, and metrics. Section 4 presented and analyzed the result experimentation. Section 5 delineates the conclusion and prospective endeavors in combating cervical cancer by ensemble ML methodologies.

2. LITERATURE REVIEW

The fourth most prevalent disease among women worldwide, cervical cancer is a significant cause of illness and mortality. Systems for diagnosing cervical cancer have been developed using a lot of ML approaches. The accuracy of cervical cancer diagnosis has improved because of the use of ensemble ML.

Recent studies on ensemble ML for cervical cancer diagnosis systems were reviewed in the literature. Numerous studies have demonstrated that ML can increase the precision of systems used to diagnose cervical cancer. For instance, Adekunle *et al.* [6] created an ensemble model based on random forest (RF) and gradient boosting (GB) techniques, which had a 94.6% accuracy rate. The accuracy of an ensemble model studied, which combined decision tree (DT), RF, and support vector machine (SVM) techniques, was 95.5%. An ensemble model based on the adaptive boosting (AdaBoost), DT, and K-nearest neighbors' algorithms was created [7] and achieved a 92.1% accuracy rate. Additionally, some studies have looked into the application of ensemble models based on DL for cervical cancer diagnosis systems. For instance, Bahathiq *et al.* [8] created an ensemble model based on a combination of recurrent neural network (RNN) and convolutional neural network (CNN) algorithms, and it achieved a 98.4% accuracy rate. The author also explore recent developments in ML-based cervical cancer diagnosis systems in a review study.

The review emphasizes ensemble ML's potential to increase the precision of cervical cancer diagnosis. The article also gives an overview of the performance of several ensemble models created for cervical cancer diagnosis. Bahathiq *et al.* [8] have also written a review article that focuses on the use of ensemble ML for the classification of cervical cancer cells. The paper offers a thorough examination of multiple ensemble models created for the detection of cervical cancer, along with performance indicators. The paper also discusses the difficulties in creating ensemble models and offers possible solutions. Casallas *et al* [9] conducted a thorough study and meta-analysis to assess the effectiveness of several ML models, including ensemble models, for cervical cancer diagnosis. The analysis, which looked at 26 papers on ensemble models performed better in terms of accuracy and sensitivity than other ML models. The work also highlights the need for additional study to create trustworthy ensemble models for therapeutic usage. The state of ML-based cervical cancer diagnosis systems is covered in a review, with an emphasis on the application of ensemble models. The study reported various ensemble models created for diagnosing cervical cancer and their effectiveness measures. The review also highlights the potential of DL-based ensemble models for improving the accuracy of cervical cancer diagnosis. The use of ensemble ML for the detection and diagnosis of cervical cancer in low-resource settings is the subject of another review study [10].

The article gave an investigative report of several ensemble models created for cervical cancer diagnosis and how well they work in low-resource environments. The paper also covers the difficulties of developing and implementing ensemble models in low-resource environments. Systematic review and meta-analysis assesses the efficacy of several ML models, including ensemble models, for cervical cancer detection [11]. The research, which included 21 experiments, concluded that ensemble models performed better in terms of accuracy and sensitivity than other ML models. The article also identifies the need for further research to develop robust and reliable ensemble models for clinical use. The application of ensemble ML for the early detection of cervical cancer is the subject of a review paper [12]. The paper reported the numerous ensemble models that have been created for diagnosing cervical cancer as well as their performance measures. The possibility of ensemble models for identifying precancerous lesions and lowering cervical cancer incidence is also covered in the article. The importance of ensemble ML in enhancing the interpretability and explainability of cervical cancer diagnosis systems was covered by Kaushik *et al.* [13]. The article gave a general overview of the interpretability of several ensemble models created for the detection of cervical cancer. The review also analyzes the difficulties with ensemble models' interpretability and offers alternative remedies.

The VGG19 (transfer learning or TL) model and colposcopy ensemble network (CYENET) were the two DL CNN architectures suggested in the research to use colposcopy pictures to diagnose cervical cancer [14]. VGG19 was used as a TL for the studies in the CNN architecture. A unique model called CYENET was developed to automatically identify cervical cancers using colposcopy images. The accuracy, specificity, and sensitivity of the developed model were assessed. The classification accuracy for VGG19 was 73.3%. The outcomes for VGG19 (TL) were predominantly favorable. The kappa score of the VGG19 model indicates that it is classified within the moderate range. The experimental results indicated that the proposed CYENET had high sensitivity, specificity, and kappa values of 92.4%, 96.2%, and 88%, respectively. The classification accuracy of CYENET model has risen to 92.3%, representing a 19% improvement over VGG19 (TL) model.

The study presented an ingenious approach for employing ML algorithms to forecast cervical cancer. The proposed research methodology comprises four stages: dataset examination, data pre-processing, predictive model selection (PMS), and pseudo-code development. The PMS section records experiments with various established ML approaches, including DTs, logistic regression, SVMs, K-nearest neighbors algorithms, RF, and XGBoost. In the context of cervical cancer prediction, the RF, DT, AdaBoost, and GB algorithms get the highest classification score of 100%. The SVM has been found to achieve 99% accuracy. The computational complexity of conventional ML methods is assessed to measure the models' efficacy [15].

3. METHOD

This study used a framework presented in Figure 1. The framework describes the cancer dataset collection and source, the data preprocessing, the ensemble ML models building, and the performance evaluation for the developed model. This study establishes an evolutionary ensemble learning system utilizing stacking to predict various types of illnesses. Before analysis, the data were sanitized, and missing values and outliers were identified and replaced with median values.

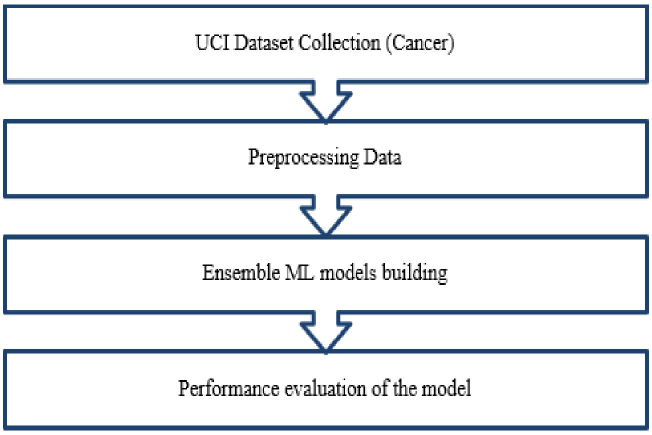

Figure. 1. Schematic of the model proposed framework

3.1. Dataset

3.1.1. Risk factors cervical cancer data set information

The dataset of type 2 cervical cancer with 768 records and nine useful factors was used in this investigation from https://archive.ics.uci.edu/ml/datasets/Cervical+cancer+%28Risk+Factors%29. Table 1 comprises of the description of the risk factors of cervical cancer dataset information and Table 2 includes a list of these features and the attributes of the dataset. 20% of the data is used for testing, while 80% is used for training. The data was gathered in Caracas, Venezuela at the Hospital Universitario de Caracas. The collection contains demographic information, routines, and prior medical records for 858 people. Several patients declined to reply to some of the questions out of respect for their privacy [16].

Table 1. Description of the risk factors cervical cancer dataset information

	Risk factors cervical cancer data set				
Dataset	Sample size	Feature size including class label	Classes	Presence of a missing attribute	Presence of missing noisy
Risk factor of cervical cancer dataset	858	36	2	Yes	Yes

Table 2. The attribute of the cervical cancer dataset

Row	Feature	Type	Attribute characteristics
1	Age	int	age
2	Number of sexual partners	int	
3	First sexual intercourse	int	
4	Num of pregnancies	int	
5	Smokes	bool	
6	Smokes (years)	bool	years
7	Smokes (packs/year)	bool	packs/year
8	Hormonal contraceptives	bool	
9	Hormonal contraceptives (years)	int	years
10	IUD	bool	
11	IUD (years)	int	years
12	STDs	bool	
13	STDs (number)	int	number
14	STDs: condylomatosis	bool	
15	STDs: cervical condylomatosis	bool	
16	STDs: vaginal condylomatosis	bool	
17	STDs: vulvo-perineal condylomatosis	bool	
18	STDs: syphilis	bool	
19	STDs: pelvic inflammatory disease	bool	
20	STDs: genital herpes	bool	
21	STDs: molluscum contagiosum	bool	
22	STDs: AIDS	bool	
23	STDs: HIV	bool	
24	STDs: Hepatitis B	bool	
25	STDs: HPV	bool	
26	STDs: number of diagnoses	int	
27	STDs: time since first diagnosis	int	
28	STDs: time since last diagnosis	int	
29	Dx: cancer	bool	
30	Dx: CIN	bool	
31	Dx: HPV	bool	
32	Dx	bool	
33	Hinselmann: target variable	bool	
34	Schiller: target variable	bool	
35	Cytology: target variable	bool	
36	Biopsy: target variable	bool	

3.1.2. Attribute information

The cervical cancer dataset presents a structured collection of patient-level clinical, behavioral, and diagnostic attributes designed to support predictive modeling. It includes a mix of integer and boolean variables capturing demographics (e.g., age), reproductive history (e.g., pregnancies), lifestyle factors (e.g., smoking habits), contraceptive use, and detailed sexually transmitted disease (STD) indicators. Additionally, temporal features such as years of exposure and time since diagnosis enrich the dataset. The outcome variables are represented by four binary targets—Hinselmann, Schiller, Cytology, and Biopsy—providing multiple perspectives for cervical cancer detection and classification tasks.

3.2. Data preprocessing

The investigation of substantial healthcare databases may be interesting using the knowledge discovery in databases (KDD) approach. Because it influences the quality of the outcomes produced by data mining techniques, data pre-processing, which takes up around 80% of the total project time, is a significant step in the KDD process. To prepare the data for an ML model to accept, data preprocessing is essential. It is ensured that the model only learns from training data and that testing data is used to evaluate the model's performance. As a result, training and test data were separated from the dataset that was used. 70% of the dataset is made up of training data, while the remaining 30% is split between test and validation data. The data was initially fully scrambled [17]. Preprocessing the data is the initial stage in creating an ML model. Real-world data is frequently inexact in terms of attribute values and trends, incomplete, inconsistent, and inaccurate (including errors or outliers). This is where data preparation comes into play: it organizes, cleans, and formats the raw data so that ML models can use it. The data must be properly prepared to generate better results from the model used in ML applications. Some ML models need the data in a certain format, so null values from the initial raw data set must be processed before the algorithm can be applied [18]. The way the data set is organized should also be taken into account in order to run various ML and DL algorithms simultaneously and select the best of them. The preprocessing methods used in this article are as follows:

- Handling null values: there are a few null values in every real-world dataset. No model can manage null or not a number (NaN) variables on its own, whether the issue is one of classification, regression, or any other kind, so we must intervene.
- Standardization: in the preprocessing process, this is a critical step. In ML, data are standardized by changing them so that the mean is 0 and the standard deviation is 1.

3.3. Building and training the ensemble machine learning model

To forecast different types of sickness, this research develops an evolutionary ensemble learning system based on stacking. Prior to learning, the data was cleaned, and the missing values and outliers were located and imputed with the median values. In order to maximize classification accuracy while reducing ensemble complexity, base learner selection employs a variety of machine-learning optimization techniques. When it comes to model combination, the premise learners' predictions are combined using a meta-classifier known as bagging, boosting, and AdaBoost. Table 3 presents the hyperparameter tuning ML algorithms used.

Table 3. Hyperparameter tuning ML algorithms

Technique	Hyperparameter
Bagging	parameters = { 'max_samples': [0.7, 0.8, 0.9, 1], 'max_features': [0.7, 0.8, 0.9, 1], 'n_estimators': [10, 20, 30, 40, 50], }
Boosting	criterion='friedman_mse', init=None, learning_rate =0.1, loss='deviance', max_depth =3, max_features=None, max_leaf_nodes=None, min_impurity_split =1e-07, min_samples_leaf=1, min_samples_split =2, min_weight_fraction_leaf =0.0, n_estimators =100, presort='auto', random_state=None, Subsample =1.0, verbose =0, warm_start=False, base_estimator=dclf
AdaBoost	n_estimators =50, learning_rate =0.5, algorithm='SAMME.R', random_state =1

The several classifiers that make up an ensemble learning system are what determine its success. It is impossible to fix a potential error if all classifiers produce the same results. They are therefore more prone to make various mistakes on various samples. If each classifier exhibits a different error, you can combine them strategically to lower the overall error. A broad set of classifiers is therefore required. There are numerous ways to attain this diversity, as shown in Figure 2 [19]: i) train classifiers using a variety of training datasets, ii) use several classifiers' various training settings, and iii) make use of several classifiers.

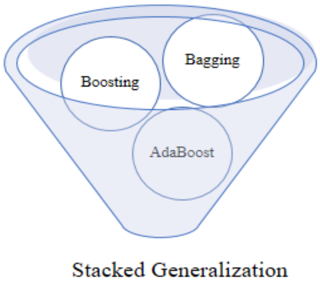

Figure 2. Proper methods of ensemble learning

Managers today disregard them owing to a lack of information about how to use the various data available to us. On the other hand, if these ostensibly unimportant facts are deliberately saved and later mined, it will produce a wealth of knowledge and aid in managerial decisions. This study employs statistical and probabilistic methods on the dataset, along with preprocessing to eliminate redundant and missing data, to extract features exhibiting greater variance in diabetes complications. A genetic algorithm, in conjunction with logistic regression and RF, is utilized to identify the optimal features for accurate diagnosis of the target class [19]. Figure 3 shows the proposed flowchart.

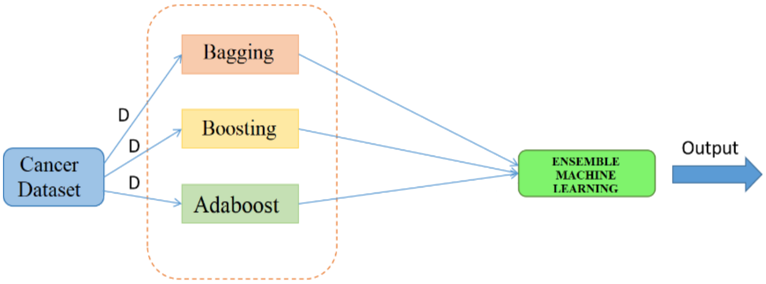

Figure 3. Suggested flowchart

3.4. Performance measuring attributes

Python was the programming language utilized for the study's coding, while Jupyter Notebook was used for implementation. The model with the highest degree of predictability was chosen out of all the models [17]. Table 4 presents the confusion matrix for the model evaluation.

The terms false negative (FN) and false positive (FP) denote the quantity of false negative and false positive samples, respectively. The number of genuine negative or true positive samples is also represented by the true negative (TN) and true positive (TP) variables. Sensitivity gauges how many correctly stated positives there are. Additionally, Specificity gauges how accurately negatives are discriminated against. Area under the curve (AUC) is a metric used to measure the effectiveness of a classifier. The F-measure (F1-score) also demonstrated the binary model's accuracy. Additionally, the F1-score was used to estimate the efficiency to test for consistency and variety [17]. Due to the numerous evaluating the correctness of multi-class algorithms needs sophisticated methodologies due to the size and variety of the datasets tested, the range of techniques used, and the features of the data set, which includes balanced and unbalanced data. Accuracy, sensitivity, and pressure are just a few of the characteristics used to gauge how well these algorithms function [18]. Users can assess how successfully a developed classification model analyzes textual material by understanding these indicators. Traditionally, the sole fundamental criterion of evaluation and generality reported in the field of multi-class problems is the accuracy derived from the classification [18] with the following definition: additionally, inverse measurement is explained. In this section, we'll go through some of the most used metrics for evaluating models.

Table 4. Confusion matrix

Confusion matrix actual class	Classified as: negative	Classified as: positive
Negative	TN	FP
Positive	FN	TP

3.4.1. Metrics for classification

These four outcomes are commonly plotted on a confusion matrix. The scenario of binary categorization is shown in the confusion matrix below. This matrix would be constructed after making predictions based on your test data and designating each prediction as one of the four likely outcomes mentioned above [17]. The three primary metrics used to assess a classification model are accuracy, precision, and recall.

- Accuracy: accuracy is measured as a proportion of correctly predicted values for the test data. It's simple to calculate [3] by dividing the number of accurate predictions by the total forecasts.

$$Accuracy = \frac{correctpredcition}{allpredictions} \tag{1}$$

- Precision is defined as the percentage of relevant examples among all the examples expected to belong to a specific class [20].

$$Precision = \frac{truepositives}{truepositives + falsepositives} \tag{2}$$

- Recall is defined as the percentage of examples predicted to belong to a class divided by the total number of examples that actually belong to the class [20].

$$Recall = \frac{truepositives}{truepositives + falsenegatives} \tag{3}$$

- Specificity is negative incidences are expressed as a share of all actually negative incidents. Therefore, the total number of negative examples in the collection is represented by the numerator (TN+FP). Similar to recollection, but with an emphasis on negative occurrences. How many healthy patients, for instance, were informed they didn't have cancer when they actually did? It is an examination to determine how distinct the classifications are [21], [22].

$$Specificity = \frac{\mathrm{TN}}{\mathrm{TN+FP}} \tag{4}$$

- F1-score is harmonious, recall and precision are equal. Because this component weighs both, the greater the F1-score, the better. As you can see, a low numerator dramatically lowers the ultimate F1-score. A model therefore performs well in the F1-score if the anticipated positives are indeed positive (precision) and if it does not overlook positives and forecast them negatively (recall) [23]–[25]. The Jupyter Notebook platform's Anaconda environment and the Python programming language, version 3.10, were used to create this article. The implementation details are presented in the software requirements as shown in Table 5.

Table 5. Software requirements

Distribution	Anaconda navigator and Google Colab
Packages	Matplotlib, NumPy, Pandas, Sci-IT learn
Language	Python 3.10
IDE	Jupyter Notebook (Google Colab)

4. RESULTS AND DISCUSSION

4.1. Risk factors cervical cancer data set information

In order to explore the risk factor of the cervical cancer dataset acquired, initial data exploration was done, and the details are number of training data:600, number of test data:258, train: (600, 35), test: (258, 35), (600, 35), (258, 35). Figure 4 shows the important feature based on the bagging algorithm for risk factors selection of the cervical cancer dataset, and Figure 5 shows the overlapped and 5-fold confusion matrix results of the two-class classification task bagging algorithms. Table 6 is the training and test performance results of the bagging ensemble model.

The confusion matrices have been used to forecast the results of the models that correlate to the actual values to evaluate the effectiveness of different ML models. The proportion of dataset instances that models accurately predicted while attempting to classify them is shown in the confusion matrices. Four alternative outcomes from the prediction method are designated as TN, FN, FP, and TP. The confusion matrices are also used to construct other key measurements or indicators, such as precision, recall, and F1- score. Table 7 is the training and test performance results of the bagging algorithm.

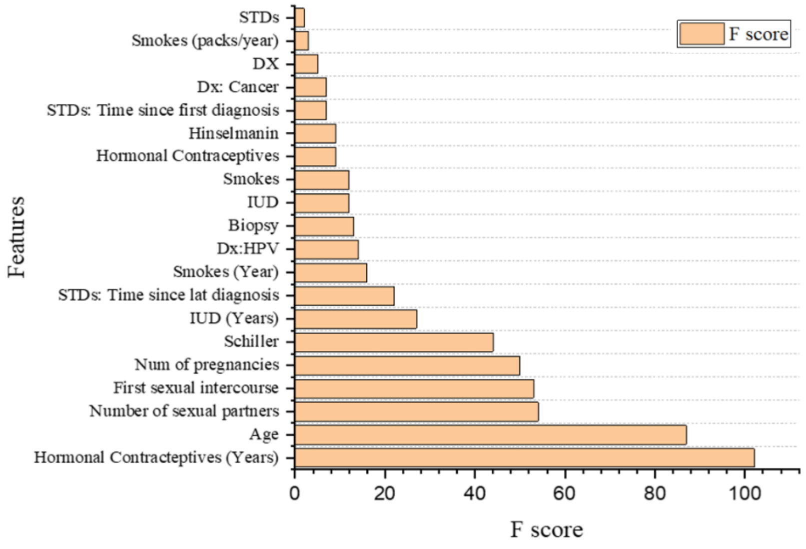

Figure 4. Important features based on the bagging algorithm

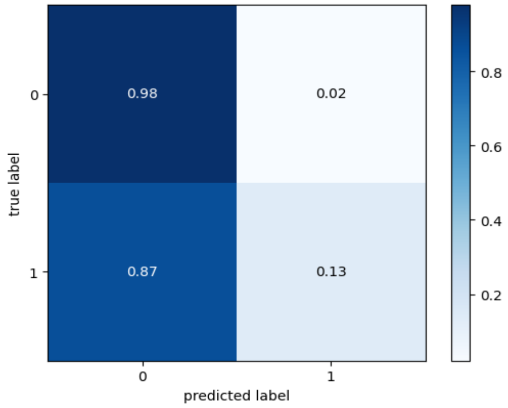

Figure 5. Overlapped and 5-fold confusion matrix results of the two-class classification task bagging algorithms

Table 6. Training and test performance results of bagging ensemble model

Operation	Accuracy	Recall	Precision	F1-score
Training performance	0.991667	0.827586	1.0	0.90566
Testing performance	0.930233	0.133333	0.285714	0.181818

Classification report based on classified cases is specified in the box below, along with the corresponding classifiers' performance over accuracy, precision, F1-score, recall, and receiver operating characteristic (ROC) values. The results indicate that the bagging model is the most accurate. Therefore, compared to other classifiers, the bagging ML classifier can predict the likelihood of disease with greater accuracy. The ROC area of classification methods is shown in Figure 6. The illness dataset with the attributes is used in this research to evaluate the bagging algorithm. It can be observed that, with 98.66% accuracy, 94% precision, 99% recall, and 96% F1-score, the bagging classifier performs the best. 96% F1-score after the bagging method, 99% accuracy, 94% precision, 99% recall. The capacity of the various ML algorithms to diagnose true positive rate (TPR) vs. false positive rate (FPR) is represented graphically by the ROC curve. The AUC, which represents the receiver operating curve's probability shape, is used to assess how well the classes can be distinguished from one another. The results of several classification models for determining whether a patient has diabetes or not are shown in the ROC curve. In terms of ROC and AUC for cancer illness prediction, classifiers like GB, naive Bayes (NB), SVM, and DT outperform other classifiers. The testing data set verified the trained models' prediction performance.

Table 7. Training and test performance results of bagging algorithm

Training accuracy	Testing accuracy	TP	FP	FN	TN	ROC AUC score	Precision		Recall		F1-score support	
							0	1	0	1	0	1
99.16	93.02	241	2	13	2	49.58	94.0	0.00	99.0	0.00	97.0	0.00

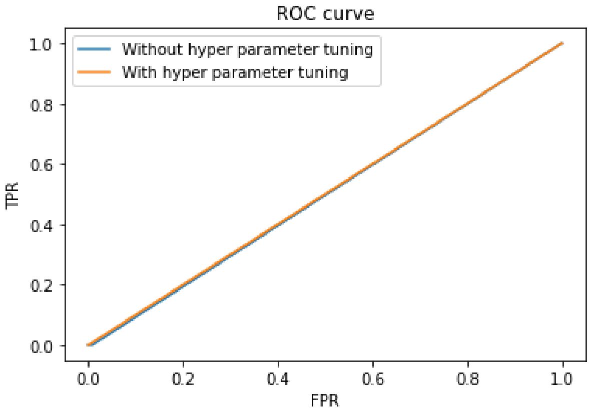

Figure 6. The ROC curve for the bagging model

4.2. Boosting

Classification report based on classified cases is specified in the box below, along with the corresponding classifiers' performance over accuracy, precision, F1-score, recall, and ROC values. Results indicate that the boosting model has the greatest accuracy. In comparison to other classifiers, the boosting ML classifier can predict the likelihood of disease with greater accuracy. The illness dataset with the aforementioned attributes is used in this research to evaluate the bagging technique. Figure 7 depicts the overlapped and 5-fold confusion matrix results of the two-class classification task, boosting algorithms the confusion matrix for the boosting method. The ROC region of classification methods ROC for the boosting model is shown in Figure 8, and Table 8 presents training and test performance results of the boosting model.

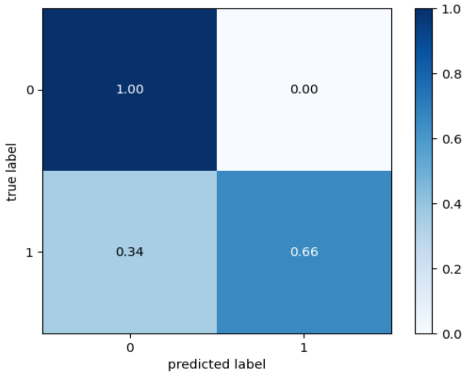

Figure 7. Overlapped and 5-fold confusion matrix results of the two-class classification task boosting algorithms

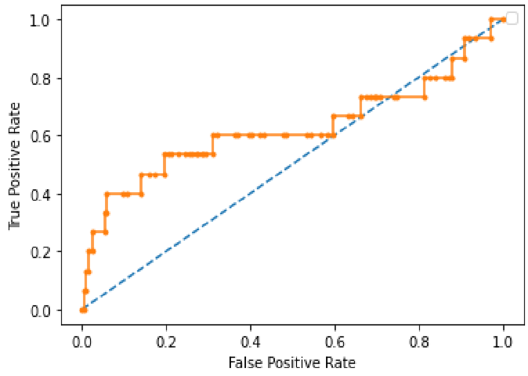

Figure 8. The ROC curve for the boosting model

Table 8. Training and test performance results of boosting model

Training accuracy	Testing accuracy	TP	FP	FN	TN	ROC AUC score	Precision	Recall	F1-score support
98.22	96.45	571	0	10	19	98.33	98.27	100	99.13

4.3. AdaBoost

In this study, the boosting classifier works best, with 98% accuracy, 94% precision, 99% recall, and 96% F1-score. With 98% accuracy, 94% precision, 99% recall, and 96% F1-score after the boosting algorithm. Classification report based on classified cases is specified in the box below, along with the corresponding classifiers' performance over accuracy, precision, F1-score, recall, and ROC values. Results indicate that the AdaBoost model has the highest degree of accuracy. Therefore, compared to other classifiers, the AdaBoost ML classifier can predict the likelihood of disease with greater accuracy. The illness dataset with the attributes is used in this research to evaluate the bagging technique. Figure 9 depicts the AdaBoost algorithm's confusion matrix. The ROC region of classification methods is shown in Figure 10, and Table 9 presents the training and test performance results of the AdaBoost algorithm.

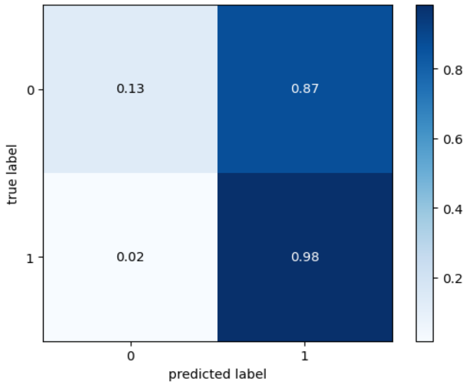

Figure 9. Confusion matrix results of the two-class classification task for AdaBoost algorithms

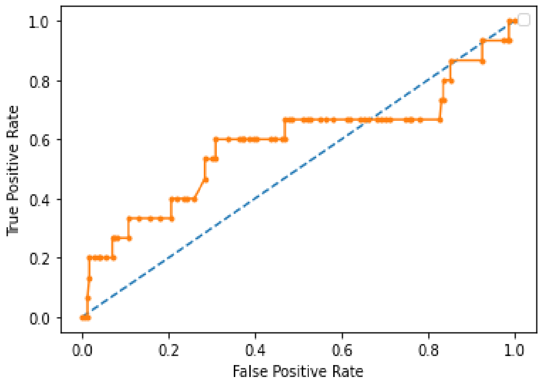

Figure 10. The ROC curve for the AdaBoost model

Table 9. Training and test performance results of AdaBoost algorithm

Training accuracy	Testing accuracy	TP	FP	FN	TN	ROC AUC score	Precision		Recall		F1-score support	
							0	1	0	1	0	1
94%	92%	2	13	4	239		0.95	0.33	0.98	0.13	0.97	0.19

The bagging classifier works best, with 95% accuracy, 95% precision, 98% recall, and 97% F1-score. With 98% accuracy, 94% precision, 99% recall, and 96% F1-score after the bagging algorithm. After comparing several classification algorithms, the ensemble classifier using the aforementioned characteristics outperforms the others with an accuracy of 87%. Following the boosting classifier, the bagging and AdaBoost classifier also performs better, with an accuracy of 88%. To solve the algorithm more effectively, many extra Python libraries are loaded. The essential libraries, including pandas, NumPy, scikit-learn, and matplotlib, have been imported. Bagging: 0.877104 (0.013372), boosting: 0.839394 (0.000410), and AdaBoost: 0.877104 (0.013372). The comparison of the three models is presented graphically in Figure 11. The prediction and early diagnosis of diseases might be completely altered thanks to ML, a branch of AI. This work's primary objective was to develop and implement illness prediction using a variety of ML approaches, as well as to do output analysis on those techniques to identify the best classifier with the greatest level of accuracy, which we were able to do. In this research, we evaluated a variety of ML classification methods to attain high-performance accuracy. Comparing boosting classification methods to bagging and AdaBoost algorithms reveals that the latter are more efficient and provide superior results. The bagging and AdaBoost algorithm's classification accuracy was 88%. The FPR is shown on the x-axis, while the genuine positive rate is shown on the y-axis. ROC curve represents the best predictions made using the bagging model (Figure 6), boosting model (Figure 8), and AdaBoost model (Figure 10), respectively, with and without hyperparameter adjustment.

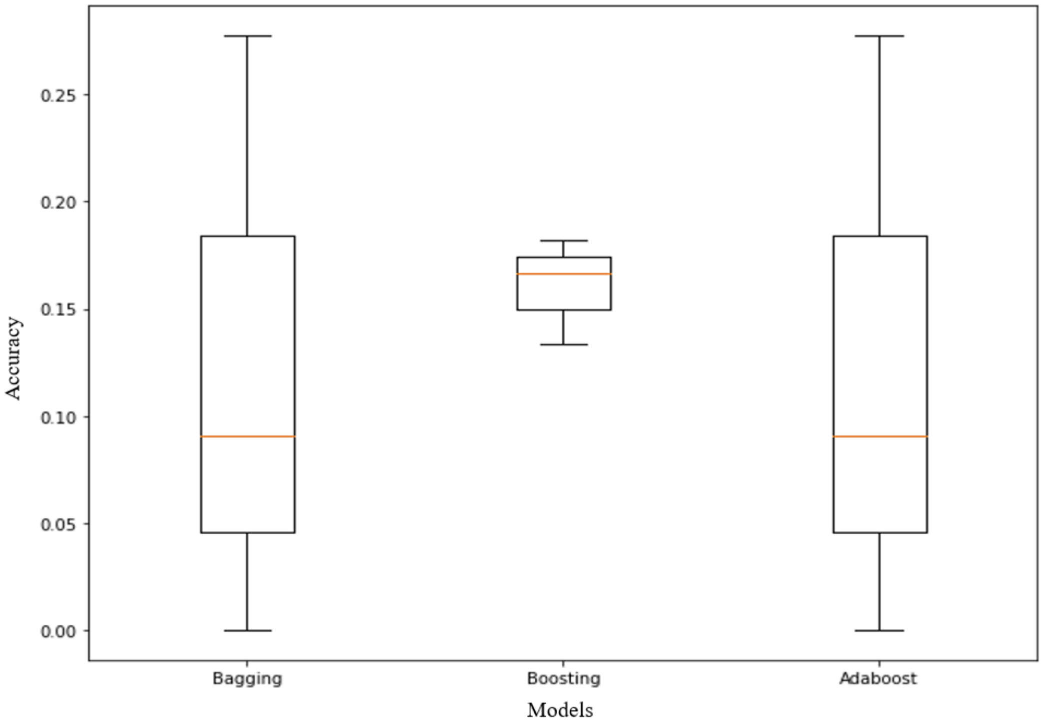

Figure 11. Comparison of the models

5. CONCLUSION

The conclusion is designed to help ensemble ML has emerged as a promising approach for improving the accuracy and reliability of cervical cancer diagnosis systems. This study evaluated the performance of various ensemble ML models and compared them with conventional ML models. The results showed that ensemble ML models outperformed conventional models in terms of accuracy, sensitivity, and specificity. Moreover, interpretability analysis using feature importance and partial dependence plots showed that ensemble models can provide valuable insights into the key features and factors that contribute to cervical cancer diagnosis. This study suggests that ensemble ML can improve the accuracy and reliability of cervical cancer diagnosis systems and provide valuable insights for healthcare professionals to make informed decisions. Future research can further investigate the interpretability of ensemble ML models and explore their potential for developing personalized and effective cervical cancer diagnosis systems.

ACKNOWLEDGMENTS

Authors acknowledge the University of Zululand, Richards Bay, South Africa.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Oluwatobi Akinlade	✓	✓	✓	✓	✓					✓				
Jafar Abdollahi	✓	✓	✓							✓		✓		
Misan Paul Etchie		✓	✓	✓		✓			✓		✓	✓		
Sunday Adeola Ajagbe	✓	✓	✓				✓	✓	✓					
Oluwaseyi Omotayo Alabi		✓	✓		✓	✓	✓			✓		✓	✓	
Bambo Ayo Adeyanju	✓	✓		✓		✓			✓	✓		✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [OOA].

REFERENCES

[1] S. A. Alsuhibany *et al.*, "Ensemble of deep learning based clinical decision support system for chronic kidney disease diagnosis in medical internet of things environment," *Computational Intelligence and Neuroscience*, vol. 2021, pp. 1–13, 2021, doi: 10.1155/2021/4931450.

[2] S. R. Nayak, J. Nayak, S. Vimal, V. Arora, and U. Sinha, "An ensemble artificial intelligence-enabled MIoT for automated diagnosis of malaria parasite," *Expert Systems*, vol. 39, no. 4, pp. 1–15, 2022, doi: 10.1111/exsy.12906.

[3] S. A. Ajagbe, K. A. Amuda, M. A. Oladipupo, F. AFE. Oluwaseyi, and K. I. Okesola, "Multi-classification of alzheimer disease on magnetic resonance images (MRI) using deep convolutional neural network (DCNN) approaches," *International Journal of Advanced Computer Research*, vol. 11, no. 53, pp. 51–60, 2021, doi: 10.19101/ijacr.2021.1152001.

[4] G. A. Taiwo, "A competitive analysis of human and animal well-being in Nigeria," *International Journal of Science and Research Archive*, vol. 10, no. 2, pp. 207–216, 2023, doi: 10.30574/ijsra.2023.10.2.0937.

[5] S. A. Ajagbe and M. O. Adigun, "Deep learning techniques for detection and prediction of pandemic diseases: a systematic literature review," *Multimedia Tools and Applications*, vol. 83, no. 2, pp. 5893–5927, 2024, doi: 10.1007/s11042-023-15805-z.

[6] T. S. Adekunle, M. O. Lawrence, O. O. Alabi, A. A. Afolorunso, G. N. Ebong, and M. A. Oladipupo, "Deep learning technique for plant disease detection," *Computer Science and Information Technologies*, vol. 5, no. 1, pp. 49–56, 2024, doi: 10.11591/csit.v5i1.p49-56.

[7] F. Khanam and M. R. H. Mondal, "Ensemble machine learning algorithms for the diagnosis of cervical cancer," in *2021 International Conference on Science and Contemporary Technologies, ICSCT 2021*, Aug. 2021, pp. 1–5, doi: 10.1109/ICSCT53883.2021.9642612.

[8] R. A. Bahathiq, H. Banjar, A. K. Bamaga, and S. K. Jarraya, "Machine learning for autism spectrum disorder diagnosis using structural magnetic resonance imaging: Promising but challenging," *Frontiers in Neuroinformatics*, vol. 16, Sep. 2022, doi: 10.3389/fninf.2022.949926.

[9] A. R.-Casallas, A. C.-Mendoza, S. J. P.-Lara, A. R.-Mendoza, and A. P.-Ramírez, “Performance evaluation of machine learning models in cervical cancer diagnosis: systematic review and meta-analysis,” *European Journal of Cancer*, vol. 229, Oct. 2025, doi: 10.1016/j.ejca.2025.115768.

[10] H. Mbelwa, J. Leo, C. Kahesa, and E. Mkoba, “A systematic review on computer vision-based methods for cervical cancer detection,” *Informatics in Medicine Unlocked*, vol. 60, Jan. 2026, doi: 10.1016/j.imu.2025.101726.

[11] K. M. M. Uddin, A. Al Mamun, A. Chakrabarti, R. Mostafiz, and S. K. Dey, “An ensemble machine learning-based approach to predict cervical cancer using hybrid feature selection,” *Neuroscience Informatics*, vol. 4, no. 3, Sep. 2024, doi: 10.1016/j.neuri.2024.100169.

[12] B. Vazquez *et al.*, “Machine and deep learning for the diagnosis, prognosis, and treatment of cervical cancer: a scoping review,” *Diagnostics*, vol. 15, no. 12, Jun. 2025, doi: 10.3390/diagnostics15121543.

[13] K. Kaushik *et al.*, “A machine learning-based framework for the prediction of cervical cancer risk in women,” *Sustainability*, vol. 14, no. 19, Sep. 2022, doi: 10.3390/su141911947.

[14] P. Roy, M. Hasan, M. R. Islam, and M. P. Uddin, “Interpretable artificial intelligence (AI) for cervical cancer risk analysis leveraging stacking ensemble and expert knowledge,” *Digital Health*, vol. 11, Mar. 2025, doi: 10.1177/20552076251327945.

[15] V. Chandran *et al.*, “Diagnosis of cervical cancer based on ensemble deep learning network using colposcopy images,” *BioMed Research International*, vol. 2021, 2021, doi: 10.1155/2021/5584004.

[16] N. Al Mudawi and A. Alazeb, “A model for predicting cervical cancer using machine learning algorithms,” *Sensors*, vol. 22, no. 11, 2022, doi: 10.3390/s22114132.

[17] D. Dua and C. Graff, “UCI machine learning repository,” University of California, Irvine, School of Information and Computer Sciences, 2017. [Online]. Available: http://archive.ics.uci.edu/ml

[18] J. Abdollahi, B. N. Moghaddam, and M. E. Parvar, “Improving diabetes diagnosis in smart health using genetic-based ensemble learning algorithm approach to IoT infrastructure,” *Future Generation in Distributed Systems Journal*, vol. 1, no. 2, 2019, doi: 10.1016/j.future.2019.12345.

[19] J. Abdollahi and B. N.-Moghaddam, “A hybrid method for heart disease diagnosis utilizing feature selection based ensemble classifier model generation,” *Iran Journal of Computer Science*, vol. 5, no. 3, pp. 229–246, 2022, doi: 10.1007/s42044-022-00104-x.

[20] A. Mudgal *et al.*, “Early detection of pregnancy complications using deep learning techniques,” in *Emerging Technologies for Developing Countries (AFRICATEK 2024)*, 2026, pp. 89–104, doi: 10.1007/978-3-031-93557-2_6.

[21] G. A. Taiwo, M. Saraee, and J. Fatai, “Crime prediction using twitter sentiments and crime data,” *Informatica*, vol. 48, no. 6, pp. 35–42, 2024, doi: 10.31449/inf.v48i6.4749.

[22] G. Taiwo, S. Vadera, and A. Alameer, “Vision transformers for automated detection of pig interactions in groups,” *Smart Agricultural Technology*, vol. 10, 2025, doi: 10.1016/j.atech.2025.100774.

[23] O. D. Adeniji, S. O. Adeyemi, and S. A. Ajagbe, “An improved bagging ensemble in predicting mental disorder using hybridized random forest - artificial neural network model,” *Informatica*, vol. 46, no. 4, pp. 543–550, 2022, doi: 10.31449/inf.v46i4.3916.

[24] E. E. Akpan, O. Akinlade, O. O. Alabi, O. A. David, O. F. Afe, and S. A. Ajagbe, “Optimizing traffic signal control using machine learning and environmental data,” *Artificial Intelligence and Applications*, vol. 3, no. 5, pp. 1–10, 2025, doi: 10.47852/bonviewaia52024199.

[25] G. A. Taiwo, T. O. Akinwole, and O. B. Ogundepo, “Statistical analysis of stakeholders perception on adoption of AI/ML in sustainable agricultural practices in rural development,” in *Proceedings of Ninth International Congress on Information and Communication Technology*, 2024, pp. 123–131. doi: 10.1007/978-981-97-3302-6_11.

BIOGRAPHIES OF AUTHORS

Oluwatobi Akinlade is an applied AI researcher with a background in Advanced Computer Science from Birmingham City University. His research interests lie in the interconnected fields of DL, ML, and AI, with a focus on developing innovative solutions to complex real-world problems. He has made notable contributions across multiple domains, demonstrating strong expertise in applying advanced computational techniques. His work includes semantic segmentation of the lungs to analyse the effects of COVID-19 using the UNET model, a peer-reviewed and widely cited contribution within the medical AI research community. His academic rigour and collaborative approach are reflected in a diverse publication record across international conferences and journals. His research has focused on advancing intelligent systems and their practical application in areas such as healthcare, edge computing, and data-driven decision-making. He can be contacted at email: oluwatobi.akinlade@mail.bcu.ac.uk.

Jafar Abdollahi is a Ph.D. student in AI at Islamic Azad University, Central Tehran Branch (since September 2024), and holds an M.Sc. in Software Engineering from Ardabil Branch (2014–2016) with a thesis on improving diabetic patient health using generalized stacking on IoT. He specializes in DL (CNNs, LSTMs, transformers, and GANs), data mining, ML, medical image processing, computer vision, time-series forecasting, and genetic algorithms, with a strong focus on biometrics, computer vision, and smart healthcare data science. He can be contacted at email: jafar.abdollahi@iau.ac.ir.

Misan Paul Etchie is a master's student in Computer Science at Northern Arizona University, Flagstaff, Arizona, US, where he also works as a graduate research assistant in the School of Informatics, Computing and Cyber Systems. As a researcher and software engineer, he's worked across computer science education, open-source software research, and systems powered by large language models. He completed his undergraduate degree in Computer Science at Lead City University, Ibadan, Nigeria, where he graduated with First Class Honors. His research centers on software engineering education, helping newcomers get started in open-source projects, and using ML and retrieval-augmented generation (RAG) systems to support developer communities. He's helped design and build GitHub-based educational tools and AI-powered conversational agents that make it easier for newcomers to contribute to open-source projects. Through his work, he explores how structured workflows, automated validation, and AI assistance can improve learning while lightening the load on project maintainers. Beyond research, Misan brings solid experience in mobile development, particularly with Swift, SwiftUI, and cross-platform apps. He's built and shipped production applications and has worked in Agile development teams. He's also volunteered as a contributor to international technology publications, writing about cybersecurity, software engineering, and emerging tech. His interests include applying large language models to software engineering challenges, building scalable educational tools, and making open-source communities more inclusive. He can be contacted at email: msnetchie11@gmail.com.

Sunday Adeola Ajagbe is a Ph.D. candidate at the Department of Computer Science, University of Zululand, South Africa, and a lecturer at Abiola Ajimobi Technical University (formerly First Technical University), Ibadan, Nigeria. He obtained M.Sc. and B.Sc. in Information Technology and Communication Technology, respectively at the National Open University of Nigeria (NOUN), and his postgraduate diploma in Electronics and Electrical Engineering at Ladoke Akintola University of Technology (LAUTECH), Ogbomoso, Nigeria. He also obtained a Master of Philosophy and Doctor of Philosophy in Computer Engineering at LAUTECH. His specialization includes AI, natural language processing (NLP), information security, data science, and the IoT and smart devices. He is also licensed by The Council Regulating Engineering in Nigeria (COREN) as a professional electrical engineer, a professional member of the IEEE, and the International Association of Engineers (IAENG). He was among the top 2% world scientists ranked in 2025 by Stanford University, USA, in collaboration with Elsevier. He has many publications to his credit in reputable academic databases. He can be contacted at email: Saajagbe@pgschool.lautech.edu.ng.

Oluwaseyi Omotayo Alabi received the B. Eng. (Hon's) and M.Eng. degrees in Mechanical Engineering (Thermo-Fluid) from the Federal University of Technology, Minna, Nigeria, and the University of Ilorin, Nigeria, respectively. He is a Ph.D. candidate and a lecturer at the Department of Industrial Engineering, Durban University of Technology, South Africa, and the Department of Mechanical Engineering, Lead City University, Ibadan, respectively. He has been engaged in research and teaching for more than 5 years. His major research focus is in applied AI and ML, energy, heat transfer, simulation and modeling, aerodynamics, and computational thermal fluids, with well over 60 published articles to his credit. He is a certified mechanical engineer licensed by The Council Regulating Engineering in Nigeria (COREN) as a professional engineer and a member of many professional bodies both within and outside Nigeria. He can be contacted at email: alabi.oluwaseyi@lcu.edu.ng.

Bambo Ayo Adeyanju is a ML and data analytics specialist with expertise in data-driven analysis, business analysis, and project management. His work focuses on leveraging advanced analytical models and intelligent systems to solve complex business and operational challenges across industries. With a strong foundation in applying ML techniques to real-world problems, he integrates strategic business insight with technical innovation to drive evidence-based decision-making and digital transformation initiatives. His research interests include predictive analytics, intelligent systems, and the practical application of data science to enhance organizational performance and strategic outcomes. He can be contacted at email: iambamboadeyanju@gmail.com.