INTERNATIONAL JOURNAL OF SOFTWARE ENGINEERING AND COMPUTER SYSTEMS
E-ISSN: 2180-0650
VOLUME 11, ISSUE 1, 2025, pp. 92 -105
DOI: https://doi.org/10.15282/ijsecs.11.1.2025.7.0139

RESEARCH ARTICLE

Enhanced stacking ensemble model for cervical cancer prediction

Fatima Isa Jibrin*, Lawan Jibrin Muhammad

Faculty of Science, Department of Computer Science, Federal University of Kashere 0182, Nigeria

Abstract - Early prediction of cervical cancer is crucial for positive treatment outcomes and reducing the high mortality rate among women. This research aims at enhancing ensemble Machine Learning model to improve existing methods and enable early diagnosis of the disease. Traditional methods have limitations in terms of sensitivity and specificity; hence, there is a need for more accurate and reliable predictive models. Existing machine learning models often face challenges such as overfitting, underfitting, and variability in prediction performance; this research, which integrates different Machine Learning techniques to improve prediction, offers a potential solution to these challenges. In this study, an ensemble approach was used to develop an enhanced ensemble Machine Learning model using Support Vector Machine (SVM), Extreme Gradient Boosting (XGB), and Random Forest (RF). The three algorithms were chosen for their robustness against overfitting and for improving prediction accuracy. Furthermore, the study overcame issues of imbalanced datasets by applying Synthetic Minority Oversampling Technique (SMOTE) and Random Undersampling (RUS) before training and testing the Machine Learning algorithms to obtain better results and improve the accuracy of the model. This study achieved a higher accuracy of 95%, 97% precision, recall of 96%, and 97% F1-score. The result indicates that the enhanced ensemble approach correctly identifies a large majority of cervical cancer cases, showing its reliability in prediction and assisting medical experts in early diagnosis. However, this study is limited to secondary datasets, which involve the collection of only secondary source datasets from the UCI repository for training and validating Machine Learning models. Future work should focus on advancing to Deep Learning to overcome the limitation of utilising feature extraction in Machine Learning algorithms and adding a lot more medical images to the dataset.

Article History
Received : 21 November 2024
Revised : 20 August 2025
Accepted : 25 August 2025
Published : 29 August 2025

Keywords
Ensemble
Machine Learning
Prediction
Cervical Cancer
Model

1. Introduction

The world has achieved a lot with improvements in innovation together with efforts in healthcare services. Machine learning models are now improving existing methods to achieve reliable results in the medical field. Cervical cancer is one of the recurring cancers related to a Sexually Transmitted Disease (STD) known as Human Papillomavirus (HP). It is an ailment whereby early prediction is very important to have a positive outcome of treatment and to reduce the rate of mortality among women. With early diagnosis of the virus that causes the disease, this virus can be eliminated at an early stage and cervical cancer becomes preventable; but when prediction is delayed, treatment becomes very expensive and complicated, thus increasing the rate of mortality among the patients [1].

Computer science as a field of study is evolving with emerging technologies, and new domains are continuously discovered, which include machine learning (ML). Recently, machine learning algorithms have been increasingly applied for predicting cancers and identifying high-risk individuals that need immediate medical attention [2]. These algorithms are very useful tools that operate by learning patterns from large datasets, and those patterns are used in making predictions. As regards cervical cancer prediction, the algorithms can learn from patients' data and identify features associated with the increased rate of the disease. Ensemble methods integrate results of individual machine learning algorithms to bring out a precise and improved prediction [3].

This research shows an ensemble approach for predicting cervical cancer by developing an enhanced ensemble machine learning model. This involves identifying risk factors associated with the disease and then collecting data and preprocessing it to address missing values and imbalanced data. After this, the pre-processed datasets are used to train the predictive models; the performance of the predictive model is analysed, and the predicted results are compared with existing results to assess the model's performance. Enhanced prediction models directly impact healthcare decision-making by correctly predicting a high proportion of actual positive cases, and having early correct predictions improves patients' treatment outcomes. Existing ensemble models for predicting cervical cancer remain suboptimal in terms of accuracy, necessitating further study to improve or enhance performance. Also, previous studies reported issues of imbalanced datasets; however, this research uses SMOTE and RUS techniques to overcome issues of imbalanced datasets to enhance accurate prediction of the existing ensemble models.

Cervical cancer has been a health concern worldwide, resulting in a high mortality rate among women. Prompt identification and precise diagnosis of the disease will lead to prevention of the disease and better outcomes of treatment. Traditional methods have limitations in terms of sensitivity and specificity [1]. Hence, there is a need for more accurate and reliable predictive models.

*CORRESPONDING AUTHOR | F.I. Jibrin | ✉ fateemaij@gmail.com
© 2025 The Authors. Published by Universiti Malaysia Pahang Al-Sultan Abdullah Press. This is an open access article under the CC BY-NC 4.0 license

Machine learning is enhancing cervical cancer prediction by leveraging large datasets and identifying complex patterns that may be missed by conventional methods. However, individual machine learning models often face challenges such as overfitting, underfitting, and variability in prediction performance. Ensemble Machine Learning, which integrates different Machine Learning techniques to improve prediction, offers a potential solution to these challenges [4]. This study's goal is to enhance an ensemble Machine Learning approach to predict cervical cancer. This model achieves higher accuracy and is more reliable than the existing methods. This research aims to enhance early detection, reduce mortality rates, and enhance better outcome of treatment.

Cervical cancer has been a health concern worldwide, resulting to high mortality rate among women. Prompt identification and precise diagnosis of the disease will lead to prevention of the disease and better outcomes of treatment. Traditional methods have limitations in terms of sensitivity and specificity [1]. Hence, there is a need for more accurate and reliable predictive models. Machine learning is enhancing cervical cancer prediction by leveraging large datasets and identifying complex patterns that may be missed by conventional methods. However, individual machine learning models often face challenges such as overfitting, underfitting, and variability in prediction performance. Ensemble Machine Learning, which integrates different Machine Learning techniques to improve prediction, offers a potential solution to these challenges [3].

This study's goal is to enhance an ensemble Machine Learning approach to predict cervical cancer. This model aims to achieve higher accuracy and to be more reliable than the existing methods. The main aspects of this study include data collection and preprocessing, training the algorithms to learn from data and testing, developing the ensemble model, and model validation and evaluation. This research aims to enhance early detection, reduce mortality rates, and enhance better outcome of treatment. This paper is organized as follows: Section 1 provides the background of the study, followed by section two which discuss about related works, section three talks about materials and methods; Section 4 contains results and discussion; and Section 5 contains the conclusion.

2. Related Works

Cervical cancer is one of the diseases with a high mortality rate globally among women. Identifying women who are most susceptible to cervical cancer will improve screening strategies to have effective prevention and also to have effective treatment of patients with the disease at the precancerous stage. Algorithms for machine learning offer promise. In predicting cervical cancer using various features that include Human Papillomavirus (HPV), weakened immune system, HIV and AIDS, smoking habits, long-term use of contraceptives, sexually transmitted diseases (STDs), and several other causes [5]. This literature review is committed to providing a summary of the current research in cervical cancer prediction using Machine Learning techniques.

Supervised Machine Learning is acquiring knowledge under supervision, like teaching a kid to walk. Until the child can walk independently, you will hold his hand, show him how to advance his foot, walk yourself as a demonstration, and so on [6]. This type of learning uses algorithms for supervised learning. Making a function or mapping out of labelled training data is the aim. An input vector (X) and an output vector (Y) make up the training data, which contains tags or labels. Each tag or label in vector Y explains its corresponding input example in vector X. These two vectors together form a training example. If there is no labelling for the input vector X, it is considered unlabelled data. In supervised learning, neural networks, decision trees, and linear regression are often utilized techniques [7].

The study in [8] suggested a model in which the performance of four classifiers was assessed to estimate the probability of eventually being diagnosed with cancer. With an AUC ranging from 0.70 one year before diagnosis to 0.97 one day prior, Random Forest yielded the best results. They conclude that although their approach is reasonable and showed good diagnostic discrimination, it had only a limited potential for discriminating earlier. Since the goal of their study was to predict diseases before they manifested, we suggest that future research look at including basic health records before a hospital referral to broaden patient histories, for instance.

Five Machine Learning algorithms, namely Decision Tree, Random Forest, Gaussian Naïve Bayes, K-Nearest Neighbor, and Logistic Regression, have all been examined in a different study by [9]. Seventy percent of the data set is meant for training, while thirty percent is meant for testing. With default parameters, Random Forest outperformed the other algorithms; Decision Tree was the least successful algorithms were GNB, KNN, and logistic regression, which came in last.

In another study by [10], the accuracy results from the ensemble approach were 1.72% higher than the average accuracy values from SVM, MLP, and KNN. When it came to sensitivity results, the ensemble method's outcomes outperformed the average of the three single classification approaches by a margin of 0.74%. The specificity results can be increased by 3.4% using the ensemble approach. The study's findings indicate that, when it comes to classifying anomalies related to Pap smear images to detect cervical cancer, the ensemble technique with the most votes can outperform the single classification method in terms of classification performance.

In study by [12], the ANN model, enhanced by SMOTE and GA, achieves 87% accuracy and 0.85 AUC, demonstrating AI's potential to lower cervical cancer mortality in Nigeria. It provides scalable solutions for public health. The study [13] demonstrates ML's potential in resource-limited settings, with bagging offering reliable predictions, though larger datasets could improve generalizability.

The performance of the existing models for predicting cervical cancer in the above literature review remains suboptimal in terms of accuracy, necessitating further study and improving ensemble machine learning models. [1] Applied K-Nearest Neighbor (KNN) and Decision Tree (DT) using a class balancing technique and suggested that other machine learning methods can be applied in the future to observe the performance of other ensemble models in prediction

and to improve the result. In this research, Support Vector Machine (SVM), Random Forest (RF), and Extreme Gradient Boosting (XGBoost) algorithms will be integrated to improve the prediction of cervical cancer. Also, previous studies reported issues of imbalanced datasets; however, this research used SMOTE and RUS techniques and has overcome issues of imbalanced datasets to obtain better results. Proper analysis of some papers reviewed is shown in Table 1.

2.1 Research Gap

The knowledge gap of previous findings and how it can be bridged: The performance of the existing models for predicting cervical cancer in the above literature review remains suboptimal in terms of accuracy, necessitating further study and improving ensemble machine learning models.

[1] applied K-Nearest Neighbor (KNN) and Decision Tree (DT) using a class balancing technique and suggested that other machine learning methods can be applied in the future to observe the performance of other ensemble models in prediction and to improve the result. In this research, Support Vector Machine (SVM), Random Forest (RF), and Extreme Gradient Boosting (XGBoost) algorithms were integrated to improve the result and accuracy of the prediction of cervical cancer. Also, previous studies reported issues of imbalanced datasets; however, this research uses SMOTE and RUS techniques to overcome issues of imbalanced datasets to obtain better results.

Table 1. Tabular representation of strengths and weaknesses of reviewed papers

Ref. Paper	Methods	Strength of the Study	Weakness
[1]	K-Nearest Neighbor, Support Vector Machine, and Random Forest Tree.	The results show that the Random Forest algorithm delivered more promising results than SVM and KNN for four target variables: Schiller, Biopsy, Hinselmann and Cytology respectively.	It was concluded that the limited amount of data, which also suffers from unbalancing brings about problems for the promising results drawn using the proposed model.
[8]	Random Forest, Complement Naive Bayes, Bernoulli Naive Bayes, and SVM	Four classifiers were evaluated for predicting a future cancer diagnosis where Random Forest achieved AUC of 0.70. This concludes that the approach is reliable.	Not extending patient histories through the integration of primary health records preceding referral to hospital.
[9]	K-Nearest Neighbor, Gaussian Naïve Bayes, Logistic Regression, Random Forest and Decision Tree (DT).	The findings demonstrate that Decision Tree algorithm was used and the study recorded an accuracy of 90%	The study did not focus on prediction duration even though the task needs to be fast
[11]	K Nearest Neighbor and Support Vector Machine	The results of their study indicate that the accuracy results obtained in the ensemble method increased by 1.72% compared to the average accuracy value in SVM and KNN.	The study focused on only two classes of abnormalities for the classification and prediction of cervical cancer.
	Decision Trees, Support Vector Machine, K-Nearest Neighbours.	The Fine Gaussian SVM The most effective classifier for categorizing Hinselmann, cytology, and biopsy The research used optimal characteristics to train various machine learning classifiers to predict cervical cancer among many risk variables.	Many datasets have been characterized by low sample size, outliers, imbalance and multiple risk factors.

Ref. Paper	Methods	Strength of the Study	Weakness
[6]	Random forest, Support vector machine and AdaBoost	The study's results included a 92.57% accuracy, 91.25% positive predictive accuracy, and 92.14% negative predictive accuracy.	Data access control, generalization, building practical applications were among the challenges. The short comings of this study include, the small sample size with the imbalanced input and target datasets.
[12]	Stacking-Integrated Machine Learning Approach, Random Forest, Logistic Regression.	This study shows that SIML enables the precise identification of high-risk females of developing cervical cancer. By maximizing the based on clinical data, behavioral patterns, and patient demographics.	The issue of imbalanced data was recorded in the study.
[13]	AdaBoost, Random forest and Support vector machine.	In this study the algorithms were integrated into many layers to generate the final stacked model. This classifier's accuracy, sensitivity, specificity, PPA, and NPA were 92.57%, 93%, 91.25%, 92.14%, and 90% respectively.	–
[14]	Decision Tree, Support Vector Machine and Multi-layer Perceptron	This study validated that machine learning can improve cervical cancer prediction. The study's findings demonstrated that the most pertinent predictors can be found using Decision Tree algorithms. Furthermore, it appears that raising patients' sociocultural standing and personal wellness may help avoid cervical cancer.	–
[15]	Random Forest, Custom Stacking Architecture	Explainable AI was utilized to improve interpretability. The model predictions were understood by the application of random forest, SHAP, and LIME.	Various issues must be addressed before deploying the classifiers in health care facilities like rigorous testing must be conducted, then comprehensive external validation.
[4]	Decision tree (DT), logistic regression (LR), support vector machine (SVM), and K-nearest	The classification score of 86% has been accomplished using decision tree (DT), and random forest (RF) algorithms.	Limited datasets were recorded.

Ref. Paper	Methods	Strength of the Study	Weakness
	neighbors (KNN).		
[16]	Logistic Regression (LR), Decision Tree (DT), Random forest (RF)	From the results obtained from this study, the voting method with SMOTE and PCA technologies helped classify cervical cancer data correctly for all target variables and raise the accuracy, sensitivity, and ROC_AUC of predictive models	In future work, the dataset used to detect cervical cancer can be improved and the efficiency of future prediction models can be increased by adding several essential attributes that support cervical cancer early detection. It could be possible to gather some data and add it to the dataset. For instance, if the HPV vaccine was administered and if the Pap smear was completed lately. To help with the development of a more accurate predictive model, this extra data can be gathered from clinics and patients. Additionally, incorporating a few key characteristics can enhance the predictive model for cervical cancer early detection.
[17]	K-Nearest Neighbor, Classification and Regression trees.	According to the study, following the training of the final prediction model verified, 87.21% of the predictions were made correctly.	–
[18]	Random forest, weighted voting ensemble, gradient boosting classifier, majority voting ensemble, k-nearest neighbor, multi-layer perceptron.	Two feature selection approaches were used to extract the most significant features. The proposed stacking architecture, applied for the first time on the cervical cancer dataset, used time elapse of 5.6 s and achieved an area under the curve score of 92.7% performing better than the methods used in previous works.	Concerning the limitations of this study, it should be reported that the size of the cancer dataset is limited and the optimal training requires a lot of malignant cases. Also, a detailed analysis of different Techniques for class balance are necessary because the danger of oversampling the minority class increases when

Ref. Paper	Methods	Strength of the Study	Weakness
			overfitting. Therefore, a thorough examination of various data is needed.
[19]	Random Forest, Support Vector Machine, K-Nearest Neighbor and Decision Tree	Results in this study show that stacking using a blend of support vector machines, random forests, and bagging were used	New hybrid classification algorithms should also be developed in the future to get more reliable classification for multiple datasets and cross-datasets.
[20]	Support Vector Machine, K-Nearest Neighbor, Decision Tree and Multilayer Perceptron	In this paper an approach using ensemble methods with SVM as the base classifier has been taken.	The accuracy obtained in this result can further be improved with nature inspired optimization algorithms and therefore will continue to exist as a potential future expansion of the work.

3. Methods and Material

This explains how listed objectives in this study will be achieved. It describes the methodology involved in the research. Also, it explains the method of data collection and its source. Dataset variables, data cleansing, enhanced models, and model validation were also discussed.

3.1 Data Collection

This is an important stage in every research and especially when it involves cervical cancer prediction. In this research, a secondary source from the UCI Machine Learning Repository was used for data collection; however, preprocessing the data is crucial to ensure the quality of the dataset.

3.1.1 Cervical Cancer Dataset

This research used a secondary dataset from the UCI Machine Learning Repository. The dataset is an electronic health record of patients. It provides valuable features that help in building robust predictive models to improve cervical cancer detection. The dataset was divided into three sets; train set, the test set, and the validation set to determine the model accuracy. The cervical cancer (factors) dataset was obtained from the UCI Machine Learning Repository website. A total of two thousand nine hundred and sixty-nine (2,969) records of patients. These features include age, pregnancies, smoking behaviors, use of contraceptives, and sexually transmitted diseases (STDs). Target variables, namely 'biopsy',' Schiller', ' Hinselmann ', and 'citology' confirm whether the person has cervical cancer or is healthy. The dataset was imbalanced; hence, oversampling and undersampling methods were employed during pre-processing.

After the data collection, the dataset was fed into the model for training, which helped the model become familiar with the dataset. Furthermore, the training dataset comprises pairs of an input vector and the corresponding result for prediction (Output), which is known as the target. After training the model, the second dataset was used for validation of the model to produce an unbiased assessment of its performance on the training dataset. This helps to identify any overfitting issues and to optimize the model's hyperparameters before it is used on unseen data. After training the model and learning from the dataset. The testing dataset gives an unbiased evaluation of the final model. So, a test dataset is fed to the model to evaluate its performance.

3.2 Architecture of the Ensemble Model

Figure 1 is the architecture of the ensemble model. It can also be explained in the following algorithm:

Algorithm 1: Architecture of the Model

Step 1: Collecting the dataset, which served as the foundation for the developed model.

Step 2: Clean and prepare a dataset for modelling by applying feature extraction.

Feature extraction involves identifying and describing the features that are relevant to cervical cancer. This step involved handling feature scaling and removing duplicate values. The Synthetic Minority Over-Sampling Technique, or SMOTE, creates instances of the minority class in order to address the problem of data imbalance. Additionally, the dataset for the Support Vector Machine model is subjected to random undersampling. RUS technique reduces the number of samples, which can be useful when SVM struggles with large datasets. The oversampling method is implemented by duplicating random data points, while undersampling is implemented by removing random data points

Step 3: The dataset is divided into two portions: typically, 70% for training and 30% for testing. This separation aids in evaluating the performance of the model on unseen data.

Step 4: Training the models (XGBoost, Random Forest, and SVM) using the pre-processed data. Each model is trained with subsets of the data due to SMOTE and random undersampling. A stacking ensemble method was used to enhance the three algorithms' robustness against overfitting so as to have high accuracy.

Step 5: Evaluate the trained models on the 30% of the dataset set aside for testing. Their performance is assessed using a confusion matrix to evaluate how well they generalize to unseen data.

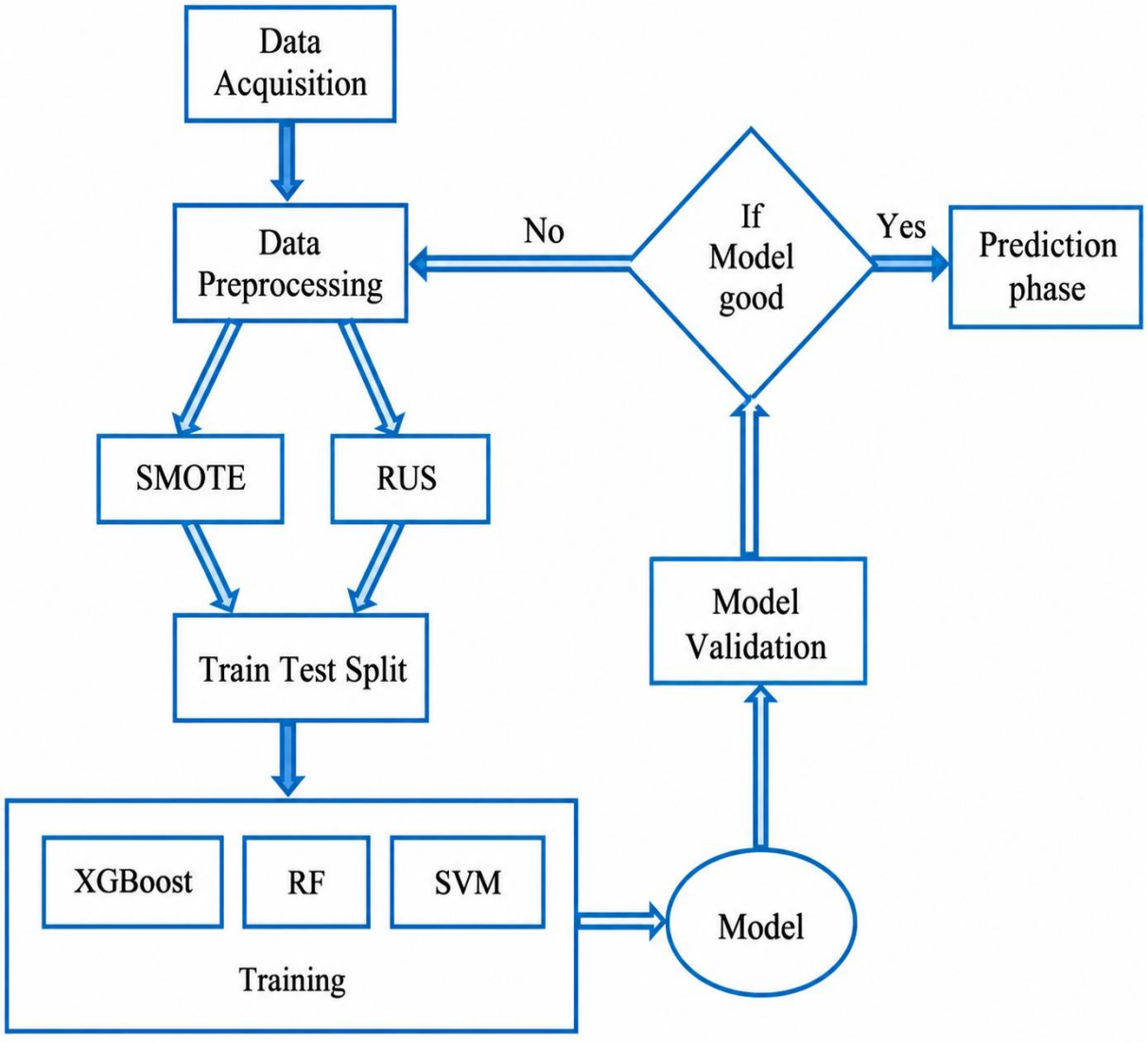


Figure 1. Architecture of the model

3.2.1 Confusion Matrix

The confusion matrix, which is composed of TP, TN, FP, and FN, is designed to provide an overview of anticipated outcomes. Several evaluation metrics are usually applied in Machine Learning and data analysis, which include the following:

3.2.2 Accuracy

Means the proportion of accurately anticipated outcomes to the total observations represented in Eq. 1.

$$Accuracy = (tp + tn)/(tp + tn + fp + fn) \quad (1)$$

True positives, or TPs, indicate that a diseased person is expected to be ill. True negatives, or healthy people correctly classified as healthy, are represented by the symbol TN. A false positive, or FP, refers to the erroneous classification of a healthy individual as ill. False negatives, or FNs, characterize a sick individual as healthy. This indicates that the most crucial component that must be as minimal as possible is FN. Another crucial component that needs to be as high as possible is TP. Hence, accuracy is important in making correct predictions of cervical cases to have effective and efficient treatment

3.2.3 Precision

It is the proportion of correctly predicted positive outcomes to the total predicted positive outcomes, represented in Eq. 2.

$$Precision = tp/(tp + fp) \quad (2)$$

Precision is important in measuring how many real positive predictions there are out of all the positive forecasts.

3.2.4 Recall

Eq. 3 shows the percentage of correctly predicted positive observations out of all observations.

$$Recall = tp/(tp + fn) \quad (3)$$

Quantifies the proportion of real positive cases that the ensemble model accurately recognized.

3.2.5 F1 Score

Represents the mean of recall and precision. The closer to 1, the better the execution or performance, as shown in Eq. 4.

$$F1\ Score = 2((precision * recall)/(precision + recall)) \quad (4)$$

The F1-score is a statistic that balances precision and recall, calculated as the harmonic mean of the two.

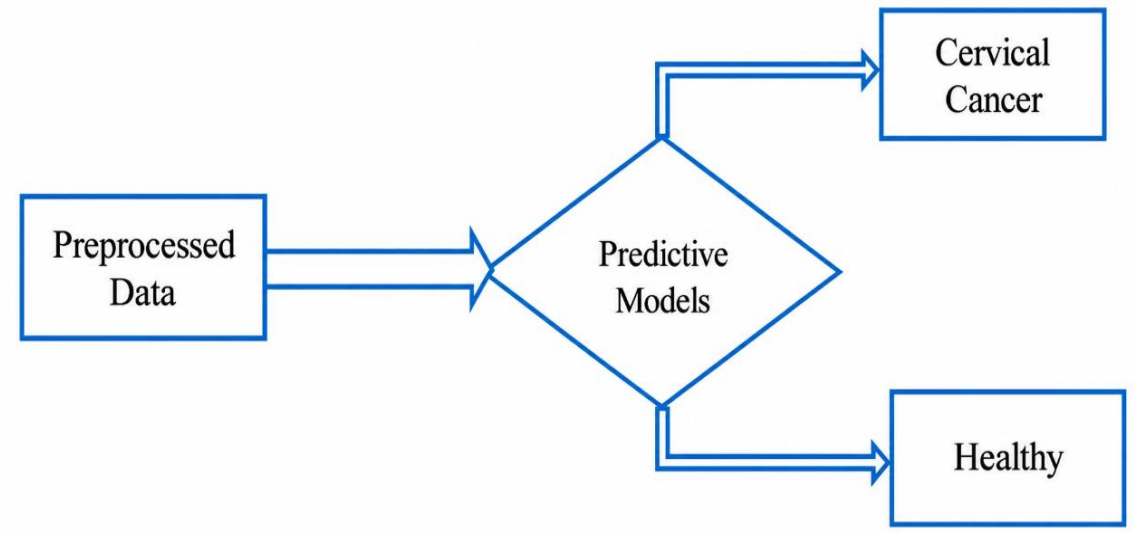


Figure 2. Prediction phase

Figure 2 is the Prediction phase. It can also be explained in the following algorithm.

Algorithm 2: Prediction Phase

Step 1: Collect data that has been preprocessed, which in this case would be the condition that you want to classify as either cervical cancer or healthy, as shown in Figure 2.

Step 2: Utilize the trained machine learning model using majority voting to predict whether the condition is cervical cancer or healthy.

Step 3: From Figure 2, the model will output predictions classifying them as either "Cervical cancer" or "healthy" based on the patterns it learned during training.

Step 4: If a model performs well and meets the criteria for training, move on to the Prediction Phase (Deployment) as shown in Figure 2. However, if a model doesn't meet the expectations, return to the data preprocessing stage, make necessary adjustments, and retrain the model until it learns effectively from the dataset.

3.2.6 Implementation

This study was implemented using the Python Programming language for the following reasons and tools. Python is a well-liked programming language for machine learning. Python's benefits include readability and simplicity. Python has a variety of strong libraries and frameworks. These libraries offer different tools for Machine Learning.

Here are the general system requirements for using Google Colaboratory (Collab):

- Device: Google Collab is accessible via a PC, laptop, tablet, or even a smartphone using a current web browser.
- Internet Connection: Google Collab cannot be accessed or used without a reliable internet connection.
- Web Browser: Google Collab performs best when using Google Chrome version 42 or later.
- The Python libraries that were used are NumPy 1.23.0, Scikit-learn, and Matplotlib 3.10

4. Results and Discussion

Table 2 and Figure 3 present the outcomes acquired from the experiment carried out on the individual machine learning algorithms and the improved ensemble machine learning algorithms.

Table 2. Results of individual algorithms and ensemble model

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Random Forest	0.94	0.72	0.85	0.78	0.93
Extreme Gradient Boosting	0.93	0.70	0.84	0.77	0.92
SVM	0.92	0.68	0.83	0.75	0.91
Ensemble Model	0.95	0.97	0.96	0.94	0.94



Figure 3. Performance of individual models and ensemble models

As shown in Table 3 and Figure 4, the improved ensemble machine learning approach has higher F1-score, ROC-AUC, accuracy, recall, and precision as presented in Table 3. This means the improved model correctly predicts a high proportion of actual positive cases (good sensitivity).

Table 3. Comparison between the ensemble and existing results

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Existing [1]	0.86	0.90	0.97	0.95	--
Ensemble	0.95	0.97	0.96	0.97	0.94



Figure 4. Comparison of performance evaluation metrics

Figure 5 displays the comparison of performance evaluation metrics of both the ensemble and existing methods. From Figure 5, it can be deduced that our improved model has performed wonderfully well by exceeding or outperforming the existing method. The model achieved an accuracy of 94% as against the 86% found in the existing literature; we achieved 97% precision compared with 90%. Likewise, for the F1-score, we have 97% compared with 95% in the literature. It was only in the recall metric that we had 96% against 97% found in the existing study.

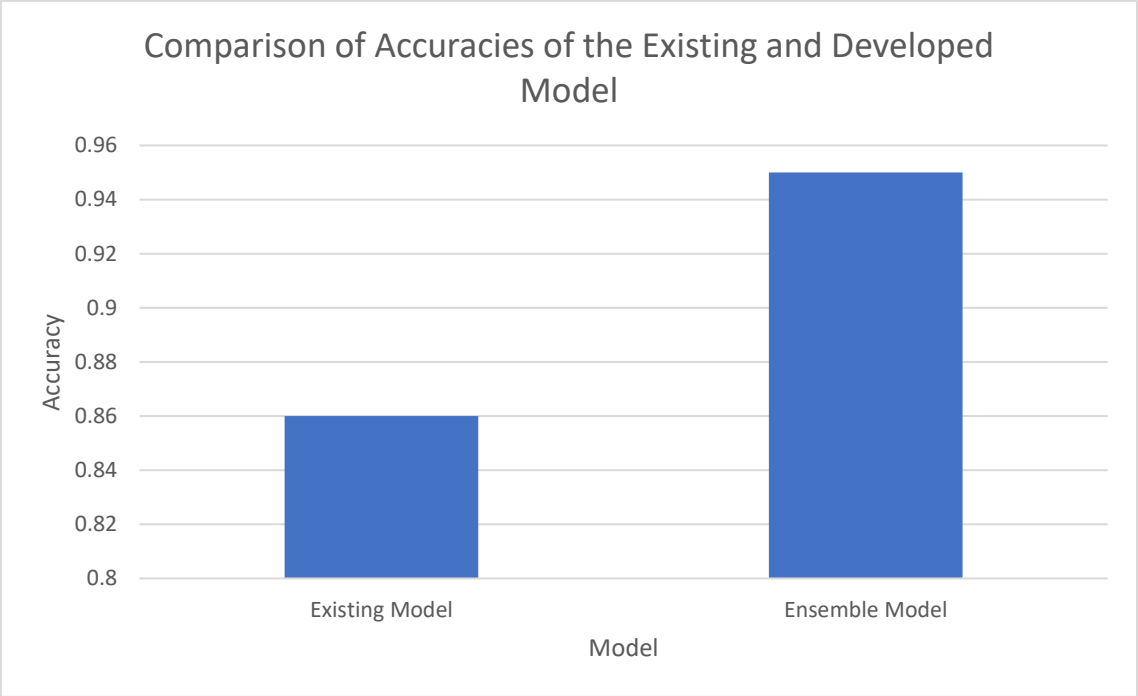


Figure 5. Comparison of accuracies of the existing and developed model

4.1 Confusion Matrix

Confusion matrices are useful for measuring the performance of classification algorithms, as they give a more detailed explanation of accuracy and errors of the model. From the confusion matrix, different evaluation metrics such as accuracy, precision, recall, and F1 score can be calculated. Figure 6 is a confusion matrix of results acquired from the ensemble Machine Learning techniques.

As presented in the confusion matrix in Figure 6, the ensemble model has correctly made 155 true positive predictions and correctly made 7 true negative predictions; however, it failed to make 4 false positive predictions and 6 false negative predictions. Hence, the ensemble model performed very well.

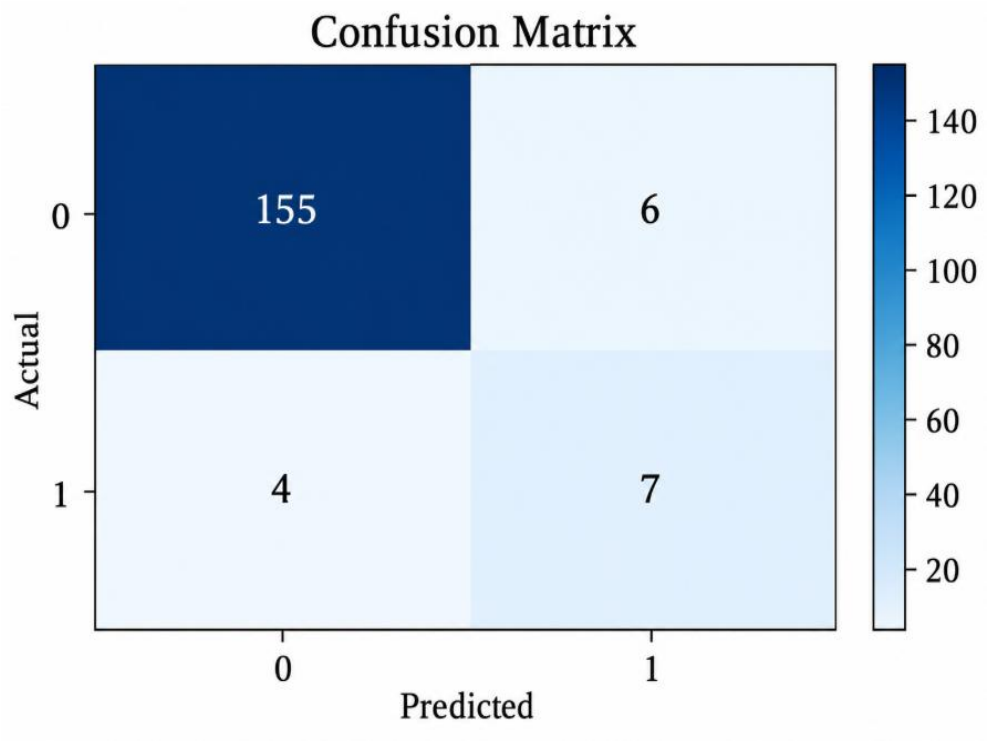


Figure 6. Confusion Matrix

4.2 Receiver Operating Characteristic (ROC) Curve

The ensemble machine learning approach's ROC-AUC is 0.94, as shown in Figure 7, indicating excellent discriminative ability. ROC-AUC evaluates the model's capacity to differentiate between positive and negative cases across different threshold values. A high ROC-AUC demonstrates that the model performs well irrespective of the decision threshold, giving a robust evaluation of the model's overall achievement.

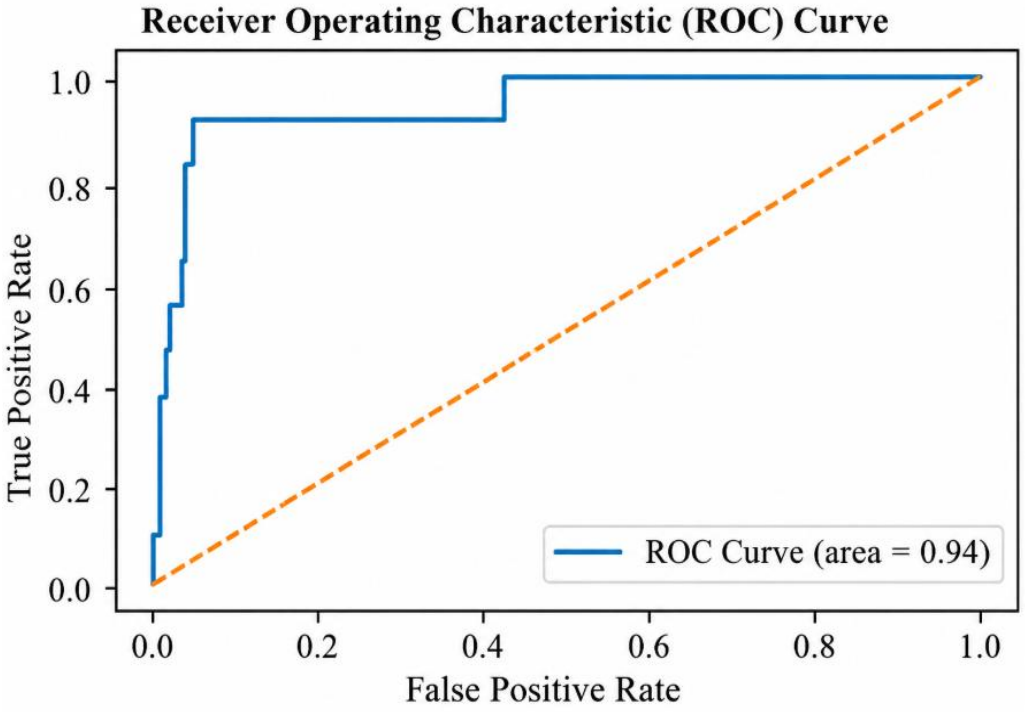


Figure 7. ROC curve

As demonstrated by the performance of the ensemble model in Figure 7, the improved ensemble approach demonstrates significant improvements in accuracy and ROC-AUC and other evaluation metrics, indicating its potential for effectively predicting cervical cancer. However, the analysis between precision and recall highlights areas for further refinement. By focusing on threshold tuning, advanced feature engineering, hyperparameter optimization, and ensuring the model's integration into clinical workflows, the model can be enhanced to provide reliable and accurate predictions.

Continuous learning and adaptation will help maintain its relevance and accuracy, ultimately contributing to better treatment outcomes and advancing the early prediction of cervical cancer. Also, higher precision and recall in the improved model indicate that it has a higher number of true positives in contrast with the existing model, which reduces unnecessary follow-up tests and potential anxiety for patients.

McNemar's Test

$$chi^2 = frac{(b - c)^2}{b + c} \quad (5)$$

This follows a chi-square distribution with 1 degree of freedom. For exact p-values (better for smaller s), we use a binomial test on the discordants, assuming under H_0 that b and c are equally likely (p=0.5). In Python (using SciPy), this is computed as `binomtest (max(b, c), s, p=0.5, alternative='two-sided').pvalue`.

1. Conservative case (least significant, largest p-value): Maximum s=35, so b=24, c=11 (b - c =13).
- Exact p-value (binomial test): ≈0.041.
- Chi-square approximation: $\chi^2 = 13^2 / 35 \approx 4.83$, $p \approx 0.028$.

2. Most significant case (smallest p-value): Minimum s=13, so b=13, c=0.
- Exact p-value: ≈0.00024.
- Chi-square: $\chi^2 = 13^2/13 = 13$, $p \approx 0.0003$.

In all possible scenarios, the p-value is <0.05 (ranging from ~0.0002 to 0.041). Therefore, we reject H_0 and conclude that the ensemble model's improvement in accuracy is statistically significant at $\alpha=0.05$.

5. Conclusions

In this study, several Machine Learning algorithms, including Random Forest (RF), Extreme Gradient Boosting (XGB), and Support Vector Machine (SVM), were used in an ensemble machine learning strategy to predict cervical cancer. The ensemble model was developed through the use of stacking. According to the analysis, the accuracy and ROC-AUC of the ensemble model were much higher than those of the previous model. In particular, the ensemble method produced results with ROC-AUC of 0.94, accuracy of 0.95, precision of 0.97, recall of 0.96, and F1-score of 0.97.

This study aimed to enhance an ensemble Machine Learning approach for the prediction of cervical cancer disease. The ensemble model combined the strengths of various algorithms to overcome the individual limitations of these algorithms so as to get higher accurate prediction of the disease. Feature extraction was also applied to address data imbalance, which also enhanced the performance of the model. The ensemble Machine Learning model demonstrates a remarkable enhancement in overall accuracy, as evidenced by the high ROC-AUC. However, the higher precision indicates an increased rate of true positives, which could reduce unnecessary follow-up tests and stress for patients. The high recall is advantageous for ensuring that most cervical cancer cases are detected, which is crucial for early intervention and treatment. The results indicate that the model is successful in categorizing positive cases, thereby enhancing the overall reliability of the predictions.

5.1 Contribution to knowledge

This study has a positive impact on machine learning and public health by providing a more reliable and interpretable model in order to predict cervical cancer early, undoubtedly contributing to better health and positive outcomes of treatment for women globally. Also, previous studies reported issues of imbalanced datasets; however, this research used feature extraction, SMOTE, and RUS techniques to overcome issues of imbalanced datasets, thereby obtaining a better result.

5.2 Recommendations

Explore more sophisticated ensemble methods like stacking to capture complex patterns in the data more effectively. Collaborate with healthcare professionals to integrate the model into clinical workflows. Implement a continuous learning framework over time, add new data to the model, and retrain it. This will guarantee that the model stays accurate and applicable as medical knowledge and patient demographics evolve.

5.3 Suggestion for Future Work

Future work should focus on advancing to Deep Learning to overcome the limitations of utilizing a feature extraction in machine learning algorithms and adding a lot more medical images to the dataset

Acknowledgements

The authors express their sincere appreciation to the management of Federal University Kashere and the Department of Computer Science for their support and encouragement. Special appreciation is extended to Gombe State University for granting study fellowship support under the Tertiary Education Trust Fund (TETFund) scheme. The authors also acknowledge the support and encouragement provided by Mr. M. B. Jibrin, Head of Department, and their families throughout the research.

Funding
This research was supported by Gombe State University through the Tertiary Education Trust Fund (TETFund) scheme under reference number GSU/REG/ACA/SS/PF/1529.

Author Contribution
Fatima Isa Jibrin: Methodology, Data curation, Resources, Writing – original draft.
Lawan Jibrin Muhammad: Supervision, Project administration, Writing – review & editing.

Conflict of Interest
The authors declare that they have no conflicts of interest.

References
[1] M. Arora, S. Dhawan, and K. Singh, "Data-driven prognosis of cervical cancer using class balancing and machine learning techniques," EAI Endorsed Transactions on Energy Web, vol. 7, no. 30, 2020. https://doi.org/10.4108/eai.13-7-2018.164264.
[2] L. J. Muhammad, E. A. Algehyne, S. S. Usman, A. Ahmad, C. Chakraborty, and I. A. Mohammed, "Supervised machine learning models for prediction of COVID-19 infection using epidemiology dataset," SN Computer Science, vol. 2, no. 1, Art. no. 11, 2021. https://doi.org/10.1007/s42979-020-00394-7.
[3] A. Tak, P. M. Parihar, S. Fatehpuriya, and Y. Singh, "Optimised feature selection and cervical cancer prediction using machine learning classification," Scripta Medica, vol. 53, no. 3, pp. 205–211, 2022.
[4] K. Kaushik, A. Bhardwaj, S. Bharany, et al., "A machine learning-based framework for the prediction of cervical cancer risk in women," Sustainability, vol. 14, no. 19, Art. no. 11947, 2022. https://doi.org/10.3390/su141911947.
[5] A. Abubakar, M. Ajuji, and I. U. Yahya, "DeepFMD: Computational analysis for malaria detection in blood-smear images using deep-learning features," Applied System Innovation, vol. 4, no. 4, Art. no. 82, 2021. https://doi.org/10.3390/asi4040082.
[6] C. H. Bhavani and A. Govardhan, "Ensemble modelling for the prediction of cervical cancer by analysing data balancing techniques," European Chemical Bulletin, vol. 12, Special Issue 3, pp. 2063–2072, 2023. https://doi.org/10.31838/ecb/2023.12.s3.259.
[7] S. Ahmed, S. Madanian, F. Mirza, and S. Zaim, "Prediction of natural gas consumption in Bahçeşehir using machine learning models," in Proceedings of the 31st Australasian Conference on Information Systems (ACIS 2020), Wellington, New Zealand, 2020.
[8] R. Weegar and K. Sundstrom, "Using machine learning for predicting cervical cancer from Swedish electronic health records by mining hierarchical representations," PLOS ONE, vol. 15, no. 8, pp. 1–19, 2020.
[9] S. El Khamlichi, I. B. A. Ouahab, M. Bouharma, F. Elouaai, A. Sedqui, and A. Muarady, "An evaluation of machine learning algorithms and feature selection methods for cervical cancer risk prediction using clinical features," International Journal of Intelligent Systems and Applications in Engineering, vol. 10, no. 4, pp. 470–479, 2022.
[10] A. Desiani, E. S. Kresnawati, M. Arhami, et al., "Majority voting as ensemble classifier for cervical cancer classification," Science and Technology Indonesia, vol. 8, no. 1, pp. 84–92, 2023. https://doi.org/10.26554/sti.2023.8.1.84-92.
[11] A. Desiani, E. S. Kresnawati, M. Arhami, et al., "Majority voting as ensemble classifier for cervical cancer classification," Science and Technology Indonesia, vol. 8, no. 1, pp. 84–92, 2023. https://doi.org/10.26554/sti.2023.8.1.84-92.
[12] L. Sun, L. Yang, X. Liu, et al., "Optimisation of cervical cancer screening: A stack integrated machine learning algorithm based on demographic, behavioral and clinical factors," Frontiers in Oncology, vol. 12, pp. 1–10, 2024.
[13] S. Jahan, M. D. S. Islam, L. Islam, et al., "Automated invasive cervical cancer disease detection at early stage through suitable machine learning model," SN Applied Sciences, vol. 3, Art. no. 806, 2021. https://doi.org/10.1007/s42452-021-04786-z.
[14] F. J. Shaikh and D. S. Rao, "Prediction of cancer disease using machine learning approach," Materials Today: Proceedings, vol. 50, no. 1, pp. 40–47, 2022.
[15] K. Chadaga, S. Prabhu, and N. Sampathila, "Predicting cervical cancer biopsy results using demographic and epidemiological parameters: A custom stacked ensemble machine learning approach," Cogent Engineering, vol. 9, no. 1, Art. no. 2143040, 2022. https://doi.org/10.1080/23311916.2022.2143040.
[16] W. Xiao, C. Liu, and X. Liu, "A hybrid LSTM-based ensemble learning approach for cervical cancer prediction," Journal of Advanced Transportation, 2021.
[17] O. Attallah, "CerCan-Net: Cervical cancer classification model via multi-layer feature ensembles of lightweight CNNs and transfer learning," Expert Systems with Applications, 2023.
[18] E. Karim and N. Neehal, "An empirical study of cervical cancer diagnosis using ensemble methods," in Proceedings of the 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), 2019.
[19] F. Khanam and M. D. R. H. Mondal, "Ensemble machine learning algorithms for the diagnosis of cervical cancer," in Proceedings of the 2021 International Conference on Science & Contemporary Technologies (ICSCT), 2021.

[20] R. Pramanik, B. Banerjee, and R. Sarkar, "MSENet: Mean and standard deviation based ensemble network for cervical cancer detection," Engineering Applications of Artificial Intelligence, vol. 123, Art. no. 106336, 2023. https://doi.org/10.1016/j.engappai.2023.106336.
[21] R. Alsmariy, G. Healy, and H. Abdelhafez, "Predicting cervical cancer using machine learning methods," International Journal of Advanced Computer Science and Applications, vol. 11, no. 7, pp. 173–184, 2020.