International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(6s) 2026 pp. 229-242
https://doi.org/10.70917/ijcisim-2026-2203
Article

Cervical Cancer Prediction Using Machine Learning with Feature Selection and Performance Evaluation

Harshitha K[1], Jagadeesha S N[2]

[1]Institute of Engineering and Technology, Srinivas University, Mukka, Mangalore, Karnataka, India.
[1]VCET, Puttur, Karnataka, India. harshithak24srinivas@gmail.com.
[2] Institute of Computer Science and Information Science, Pandeshwar, Srinivas University Mangalore, Karnataka, India. jagadeesha2012@gmail.com.

Abstract: Introduction: The issue of cervical cancer has been one of the most eminent health issues of women all over the world, particularly in the developing world where early screening and diagnosis is not very prevalent. The early diagnosis is significant in improving the survival and mortality.

Objectives: This study introduces an in-depth machine learning-based model of cervical cancer prediction based on clinical, behavioral, and demographic information.

Methods: The dataset utilized in the research is publicly available and includes 835 records of patients and 34 features, and the variable of interest was the biopsy results. This data is highly pre-processed: it is dealing with missing values, Synthetic Minority Over-sampling Technique (SMOTE) to correct the problem of a class imbalance, and Interquartile Range (IQR) to eliminate outliers. Different machine learning and deep learning algorithms like Random Forest, K-Nearest Neighbors (KNN), XGBoost, Artificial Neural Networks (ANN), and Deep Neural Networks (DNN) are used and contrasted. To improve the interpretability and efficiency of models, the process of feature selection is carried out with the help of permutation importance.

Results: The experimental findings show that the Random Forest model with the selected features has the best accuracy of 97.74% which is higher than other models with being robust and interpretable.

Conclusions: This paper demonstrates the usefulness of machine learning algorithms to predict cervical cancer at its initial stages and how it may be used to enhance healthcare decision-making.

Keywords: Cervical Cancer, Machine Learning, Random Forest, SMOTE, Feature Selection, Healthcare Analytics, Predictive Modeling.

1. Introduction

The purpose of the study is to attain a number of significant goals that are associated with the creation of an accurate and reliable cervical cancer prediction system with machine learning techniques. The first goal will be to conduct thorough Exploratory Data Analysis (EDA) to gain insight into the relationships, structure, and distribution of the clinical and behavioral information in the data set. The goal of EDA is to recognize patterns, correlations and trends that lead to a higher risk of cervical cancer. Another critical aim is to tackle the common issues faced in medical real-world data, like missing data, class imbalance, and outliers. Handling of missing values was done by using suitable imputation techniques, and class imbalance was resolved by using the synthetic minority over-sampling technique (SMOTE) to balance the positive and negative classes. The detection and handling of outliers were done with statistical methods to enhance the stability and reliability of the models. In addition, the study is designed to benchmark different machine learning (Random Forest, KNN, XGBoost, ANN) and deep learning (DNN) algorithms to find the most appropriate predictive method for cervical cancer diagnosis.

Another crucial objective of this study is to use the feature selection technique to determine the most crucial risk factors that lead to the prediction of cervical cancer. Age, smoking history, number of sexual partners, hormonal contraceptive use and IUD use were identified as important variables for prediction accuracy. According to the results obtained from this study, ensemble learning techniques like the Random Forest with selected features are far better than other models. The Random Forest model demonstrated strong performance with high accuracy, precision, specificity, and an F1 score, along with a low prediction error, suggesting its robustness and reliability for healthcare applications. The outcomes validate that feature selection is important for boosting model efficiency and clarity by eliminating less informative features. The proposed framework reinforces the potential of machine learning to aid in early detection of cervical cancer and its applications in practical clinical decision support systems, particularly in the context of limited resources.

2. Literature Survey

The recent development of machine learning had a significant impact on the field of medical diagnosis, particularly, predicting the cervical cancer. Numerous studies have been carried out to explore data-driven solutions to improve early detection, both using structured clinical data and using imaging data. More complex machine learning and deep learning models have gradually been introduced or replaced with more conventional statistical techniques such as logistic regression and decision trees since they are able to capture complex, nonlinear relationships between risk factors.

A number of research works have indicated the effectiveness of ensemble learning methods in prediction of cervical cancer. Specifically, in one study, [4], and in another, [5], it was pointed out that the Classifier of the Random Forest and Gradient Boosting are more efficient and robust than the traditional classifiers. Correspondingly, [6] and [7] demonstrated that ensemble models are efficient in working with large-dimensional medical data and decreasing overfitting with the help of aggregation mechanisms. These models are especially the ones that are applicable to structured datasets in which interaction of features is important. In recent years, deep learning approaches have become a focus as well. [8] are such studies where the authors studied Artificial Neural Networks (ANN) and Deep Neural Networks (DNN) in predicting cervical cancer.

Although these models showed good performance, they showed more benefits in large scale or image-based data as opposed to smaller structured datasets. A study by [9] also reported that deep learning models frequently entail a lot of hyperparameter tuning and computation time, and may not always warrant their small margin improvements over ensemble models in a tabular dataset setting. One of the most significant issues that are tackled in most literature is the problem of class imbalance that is prevalent in medical datasets with negative cases being far more than positive cases [10]. Other techniques that have been used like Synthetic Minority Over-sampling Technique (SMOTE) have gained extensive use. Early works by Chawla et al. and more recent ones, such as [10] and [11], showed that SMOTE enhances the sensitivity and recall of minorities in the model. Further improvement of classification has also been suggested to use hybrid methods with SMOTE and Tomek Links or Undersampling methods. Another important area of recent research has been in feature selection. High-dimensional data sets may also have redundant or irrelevant features which may hinder model performance. Recursive Feature Elimination (RFE), Principal Component Analysis (PCA) and permutation importance are among the popular methods. Indicatively, [12] demonstrated that not only does feature selection lead to increased accuracy, but also improved interpretability, which is important in a healthcare application. The age, smoking habits, and sexual behaviour variables are important risk factors that help clinicians in understanding better the predictions in the model. Besides structured data, a number of studies have been conducted on hybrid methods between clinical and imaging data. Convolutional neural networks (CNNs) were used by [13] to combine Pap smear images analysis with patient data with promising results. Nevertheless, these methods need the availability of high-quality imaging datasets, which are not always available in low-resource environments.

Comparative studies have been carried out to compare various algorithms on the same dataset. As an example, [14] pitted K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Random Forest, and XGBoost against each other, finding that ensemble models outperform them all. Although simple and interpretable, it was discovered that KNN was susceptible to noise and feature scaling. Good results were obtained with Support Vector Machines, though they need to be carefully selected with kernel and tuning of parameters. The importance of the model evaluation techniques has also been pointed out in the new works as well. Focus [15] on using stratified k-fold cross-validation to ensure believable and unbiased approximations of performance. Precision, recall, F1-score and area under the ROC curve (AUC-ROC) are becoming common metrics of evaluation, rather than accuracy alone, particularly in imbalanced data. Moreover, interpretability and explainability have already become important elements of contemporary AI-based healthcare systems. The studies of [16] (LIME) and (SHAP) highlighted the necessity of clear models that clinicians would be able to rely on, and such models were used in cervical cancer studies such as [17].

Random Forest is an ensemble model that is often favoured because of the interpretability it can have over deep neural networks. As a whole, the literature indicates a distinct tendency toward the application of ensemble learning, powerful methods of preprocessing, and feature selection to predict cervical cancer. Although deep learning models are promising, particularly in image-based diagnostic, traditional machine learning models are still very effective when dealing with structured datasets. This study expands on these results by adding SMOTE to balance classes, permutation-based feature selection, and a thorough comparison and evaluation of several models, which eventually points to the most effective and interpretable method of predicting cervical cancer.

3. Methodology

This paper has used various machine learning and deep learning models to evaluate the predictive performance and identify the most relevant model to predict cervical cancer. Neither of the models is another learning mechanism and allows one to consider the algorithms with the various strengths in the work with structured medical data. The training of the models was done on the fully preprocessed dataset which involved the unbalanced classes through SMOTE and the outliers. The hyperparameter optimization and validation procedures were also used to observe that all models were doing their best and could equally perform well under the same experimental conditions.

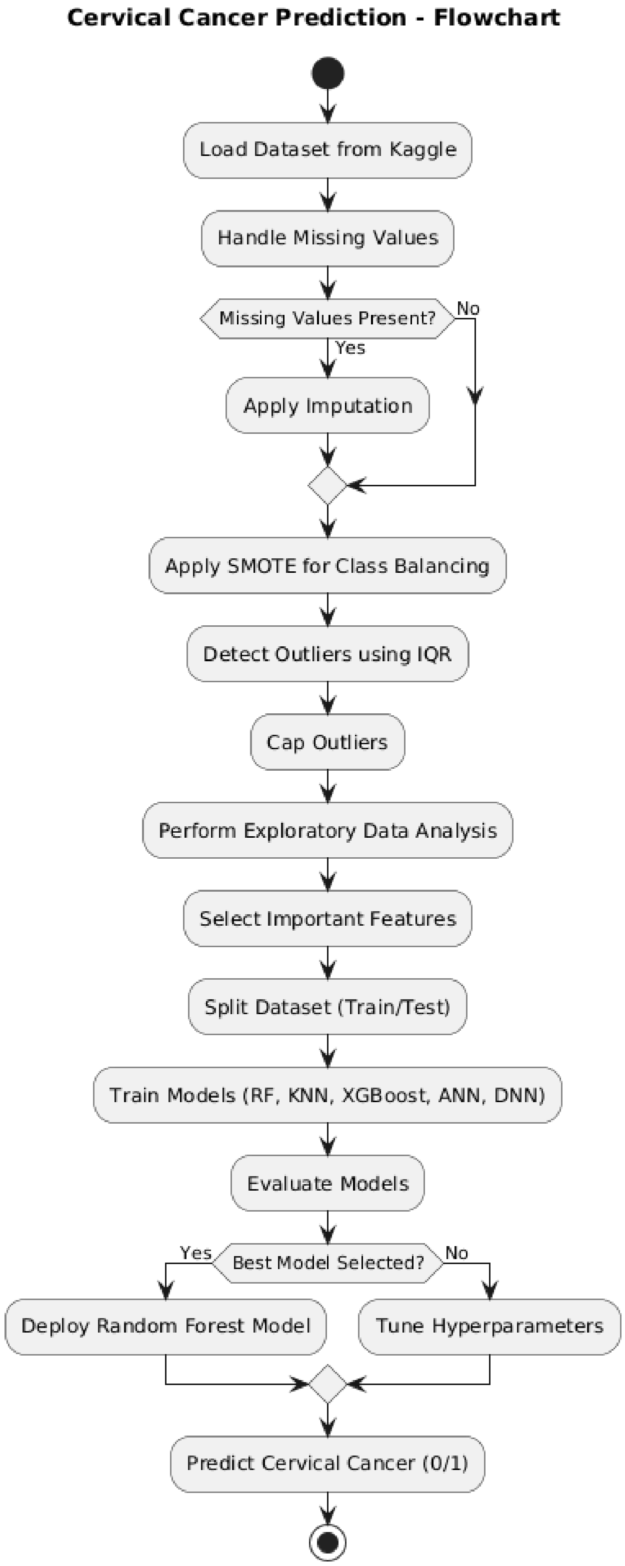


Figure 1: Flowchart for cervical cancer prediction

The presented methodology is based on a systematized pipeline of predicting cervical cancer, which entails data preprocessing [18], feature engineering, machine learning model, and evaluation. The aim is to convert raw medical data to a precise and predictable medical system. To avoid the problems of missing values, imbalance in classes and redundant features, every step is carefully designed with a perspective of the high-quality model performance.

3.1 Dataset Representation

The dataset that was utilized in the present research was found at Kaggle and includes 835 records of patients with 34 features. They are all patients, characterized by both clinical and behavioral differences, as well as demographic characteristics [19]. The dataset can be mathematically expressed as:

$$D = \{(x_i, y_i)\}_{i=1}^{n}$$

where $x_i \in R^m$ represents the feature vector with $m = 33$ attributes, and $y_i \in \{0, 1\}$ denotes the target variable (Biopsy result). The objective is to learn a mapping function:

$$f(x) \rightarrow y$$

that predicts whether a patient has cervical cancer.

3.2 Data Preprocessing

Preprocessing of data guarantees that the data is clean, consistent and appropriate to be fed into machine learning algorithms.

3.2.1 Missing Value Handling

The data has missing values which are denoted by (?). They were substituted with statistical imputation methods like mean or median substitution [20]. Missing values that were associated with a feature X were substituted as:

$$X_{impute}{}^{d} = median\ (X)$$

Skewed distributions were handled with median imputation to ensure that the distribution is robust.

3.2.2 Class Imbalance Handling

Medical datasets have a tendency to be imbalanced in terms of classes, with many negative cases considerably outnumbering positive cases. To deal with this, the Synthetic Minority Over-sampling Technique (SMOTE) was used.

SMOTE produces artificial samples by interpolating between the instances of minority classes:

$$x_{ne}{}^{W} = x_i + \lambda(x_{nn} - x_i)$$

where x^{nn} is a nearest neighbour of x_i, and $\lambda \in [0, 1]$. This procedure equalizes the dataset such that there is equal representation of the two classes and the model sensitivity is enhanced.

3.2.3 Outlier Detection and Treatment

Outliers were detected using the Interquartile Range (IQR) method:

$$IQR = Q_3 - Q_1$$

$$Lower\ Bound = Q_1 - 1.5 \times IQR$$

$$Upper\ Bound = Q_3 + 1.5 \times IQR$$

Capping was used to put limits on values that were not within these limits [21], rather than eliminating them. The method conserves infrequent, yet clinically valuable instances and stabilizes model training.

3.3 Exploratory Data Analysis (EDA)

EDA was performed to learn about the feature's distributions, relationship, and patterns. The statistical summaries and visualization indicated that behavioral variables like smoking and number of sexual partners are variable among the cancer-positive cases.

Linear relationships between features were measured using the correlation analysis:

$$r = \Sigma(x_i - \overline{x})(y_i - \overline{y}) / (\sigma_x\ \sigma_Y)$$

The majority of the features were weakly correlated which means that there was little multicollinearity and that many independent predictors could be included [22]. Skewed distributions were also discovered using EDA and were used to make preprocessing decisions like scaling and feature selection.

3.4 Feature Selection

The feature selection decreases the number of dimensions and enhances the interpretability of the models. Permutation Feature Importance was applied in this research.

The significance of a feature is quantified as the drop in model performance over shuffling the feature:

$$FI_j = Accuracy_{original} - Accuracy_{shuffled}(j)$$

Those features that had high importance scores were kept. The last features that were chosen were the age, number of sexual partners, duration of smoking and use of contraceptives [23]. This step not only minimizes noise, but also increases computational efficiency, and generalization.

3.5 Model Implementation

Various machine learning and deep learning models were deployed to learn various learning patterns.

Random Forest

A combination of decision trees with a final prediction being obtained through majority voting:

$$y = mode(T_1(x), T_2(x), ..., T_n(x))$$

Random Forest is effective in dealing with nonlinear relationships and gives feature importance, which makes it useful in medical datasets [24].

K-Nearest Neighbors (KNN)

KNN classifies a sample, using its closest neighbours:

$$y = majority\ (y_{k-nearest})$$

Euclidean distance is usually used to measure distance:

$$d(x_i, x_j) = \sqrt{}\Sigma(x_i - x_j)^2$$

KNN captures local patterns but is sensitive to scaling and noise.

XGBoost

XGBoost is used to construct a model sequentially and minimize a loss function:

$$L = \Sigma l(y_i, \hat{y}_i) + \Omega(f)$$

where l is the loss function and Ω is a regularization term to avoid overfitting.

Artificial Neural Network (ANN)

ANN calculates the output based on the weighted sums and the activation functions:

$$z = \Sigma\, w_i\, x_i + b$$

$$\alpha = \sigma(z)$$

with σ an activation function like ReLU or sigmoid.

Deep Neural Network (DNN)

DNN expands on ANN with multiple hidden layers [25], allowing the learning of features in a hierarchical manner. Preventive overfitting by regularization techniques (dropout) were employed.

3.6 Model Evaluation and Testing

The standard classification metrics were used to test the model performance:

Accuracy

$$Accuracy = (TP + TN) / (TP + TN + FP + FN)$$

Precision

$$Precision = TP / (TP + FP)$$

Recall

$$Recall = TP / (TP + FN)$$

F1-Score

$$F1 = 2 \times (Precision \times Recall) / (Precision + Recall)$$

Stratified k-fold cross-validation was used to achieve robustness. The idea is to divide the data into k subsets, maintaining the distribution of classes in an accurate manner to estimate the performance.

3.7 Algorithmic Workflow (Pseudocode)

```
Input: Dataset D | Output: Trained prediction model
1. Load dataset D
2. Handle missing values using imputation
3. Apply SMOTE to balance classes
4. Detect and cap outliers using IQR
5. Perform Exploratory Data Analysis
6. Select important features using permutation importance
7. Split dataset into training and testing sets
8. Train models: RF, KNN, XGBoost, ANN, DNN
9. Evaluate models using accuracy, precision, recall, F1-score
10. Apply cross-validation
11. Select best-performing model (Random Forest)
12. Output final prediction model
```

3.8 Final Model Selection

The final model was selected as the Random Forest model with the selected features based on the results of the evaluation. It obtained the best accuracy and was also able to perform stably across validation folds [26]. The model offers a trade-off between:

- High predictive accuracy
- Interpretability
- Computational efficiency

4. Results

This research project has proven that the different machine learning and deep learning models are effective in cervical cancer prediction based on clinical and behavioral data. Following the use of the preprocessing methods like the treatment of missing data, the balancing of classes, and the treatment of outliers, all the models demonstrated good performance in various evaluation measures [27]. Accuracy, sensitivity, specificity, precision, F1-score, and RMSE were used to compare the reliability of models. These results indicate that the ensemble-based models, especially the Random Forest and XGBoost models are always better than the other models. Also, feature selection also increases model performance resulting in better accuracy and a smaller prediction error. On the whole, the findings indicate the significance of choosing the right algorithms and characteristics to make medical diagnosis reliable.

Table 1: Comparison of different algorithms with different metrics

Algorithm	Accuracy	Sensitivity	Specificity	Precision	Recall	F1-Score	RMSE
Random Forest (All Features)	0.9712	0.9615	0.9809	0.9804	0.9615	0.9709	0.16957
Random Forest (Selected Features)	**0.9744**	**0.9615**	**0.9873**	**0.9868**	**0.9615**	**0.9740**	**0.15987**

KNN	0.9489	0.9872	0.9108	0.9167	0.9872	0.9506	0.22609
XGBoost	0.9617	0.9423	0.9809	0.9800	0.9423	0.9608	0.19580
ANN	0.9233	0.8846	0.9618	0.9583	0.8846	0.9200	0.27691
DNN	0.9265	0.9038	0.9490	0.9463	0.9038	0.9246	0.27108

Table 1 includes an overall performance analysis of different machine learning and deep learning models in cervical cancer prediction with several evaluation measures. Random Forest with selected features has the highest accuracy and specificity of all the models, which means that the overall performance is the best, and the model is more likely to correctly classify a negative case [28]. It is also very precise and F1-score, which indicates balanced and faithful classification. The model of Random Forest (all features) has a similar performance as shown in Figure 2, but the result is slightly worse, which once again proves that feature selection improves the model efficiency and accuracy. XGBoost is also a competitive alternative because it has a high specificity and precision.

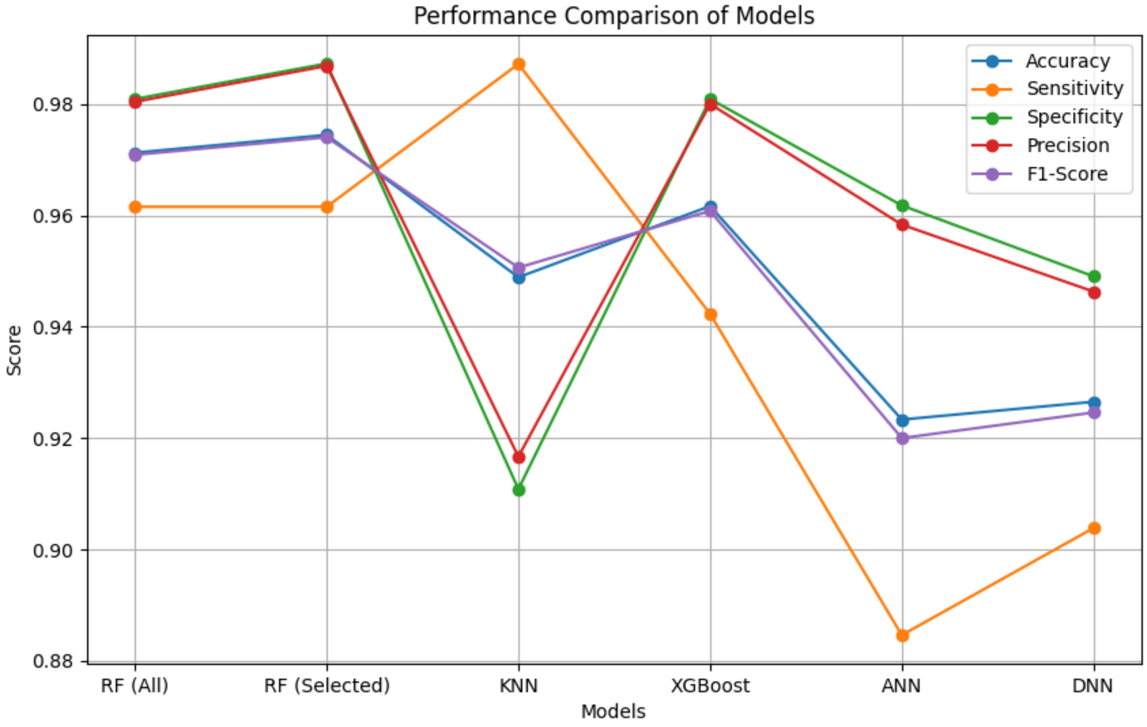


Figure 2: Performance Comparison of Models

KNN model is the most sensitive and recalls the highest number of positive cases but its low specificity implies that it has more false positives. ANN and DNN, in turn, have relatively poorer performance in most metrics, which is probably explained by the small scale of the dataset. Also, the lowest RMSE values are obtained with Random Forest models, which means that they are more predictable and have low error.

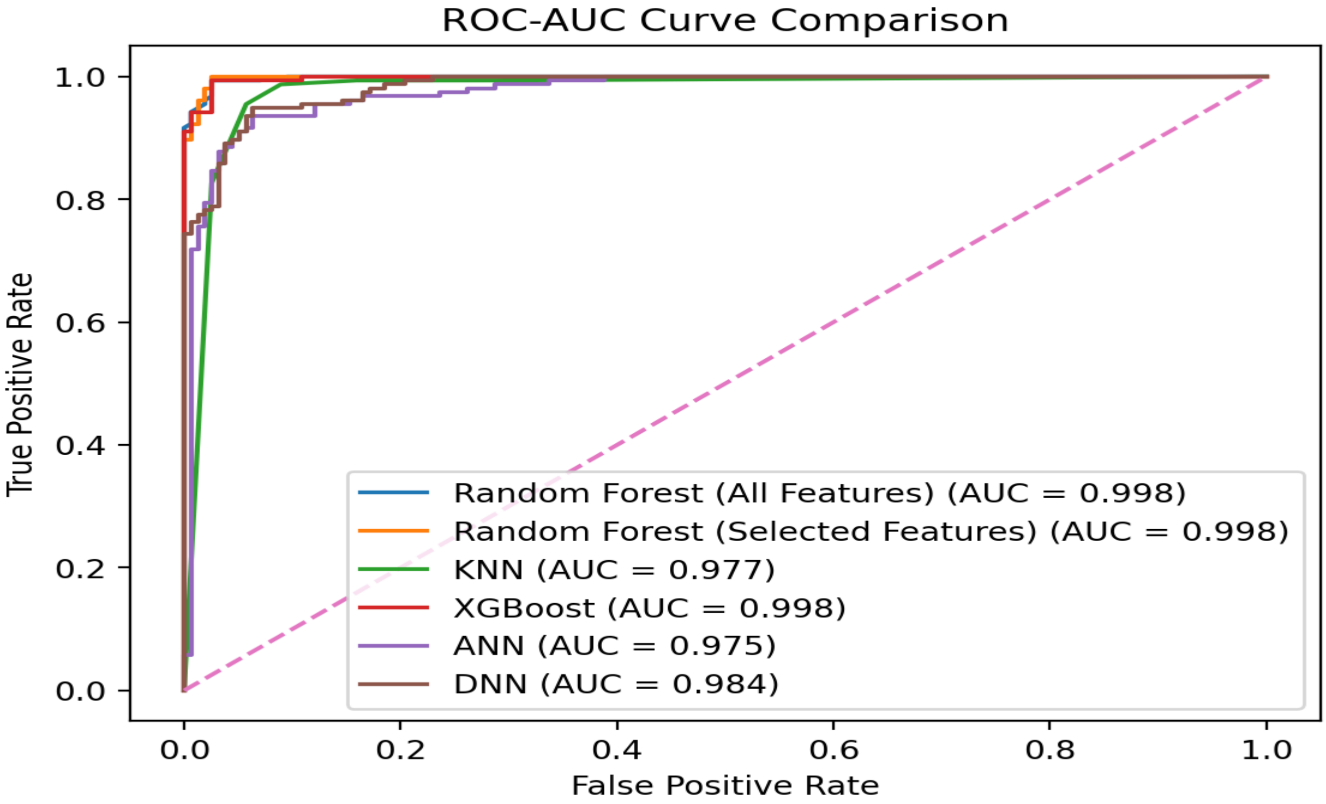


Figure 3: ROC-AUC Curve Comparison

Figure 3 presents the ROC-AUC curve, which compared the results of several machine learning and deep learning models utilized in predicting cervical cancer. The Receiver Operating Characteristic (ROC) curve shows the True Positive Rate (Sensitivity) vs the False Positive rate (1 - Specificity) at different threshold points. The diagonal dotted line depicts a random classifier. Based on the graph, it is clear that all models are much better than the baseline, since they are placed near the top-left corner.

The XGBoost and the Random Forest models (full features and selected features) have the highest Area Under the Curve (AUC) of 0.998 which means close to perfect classification performance. This indicates that the models are very useful in the differentiation of cancer-positive and cancer-negative cases. Deep Neural Network (DNN) performs well too with an AUC of 0.984, then KNN (0.977) and ANN (0.975). Even though they are a bit lower, the values also suggest a high predictive power. The graph shows that the ensemble models, especially the Random Forest and XGBoost models are more effective than the other methods and hence, the most reliable models to use in this medical classification problem.

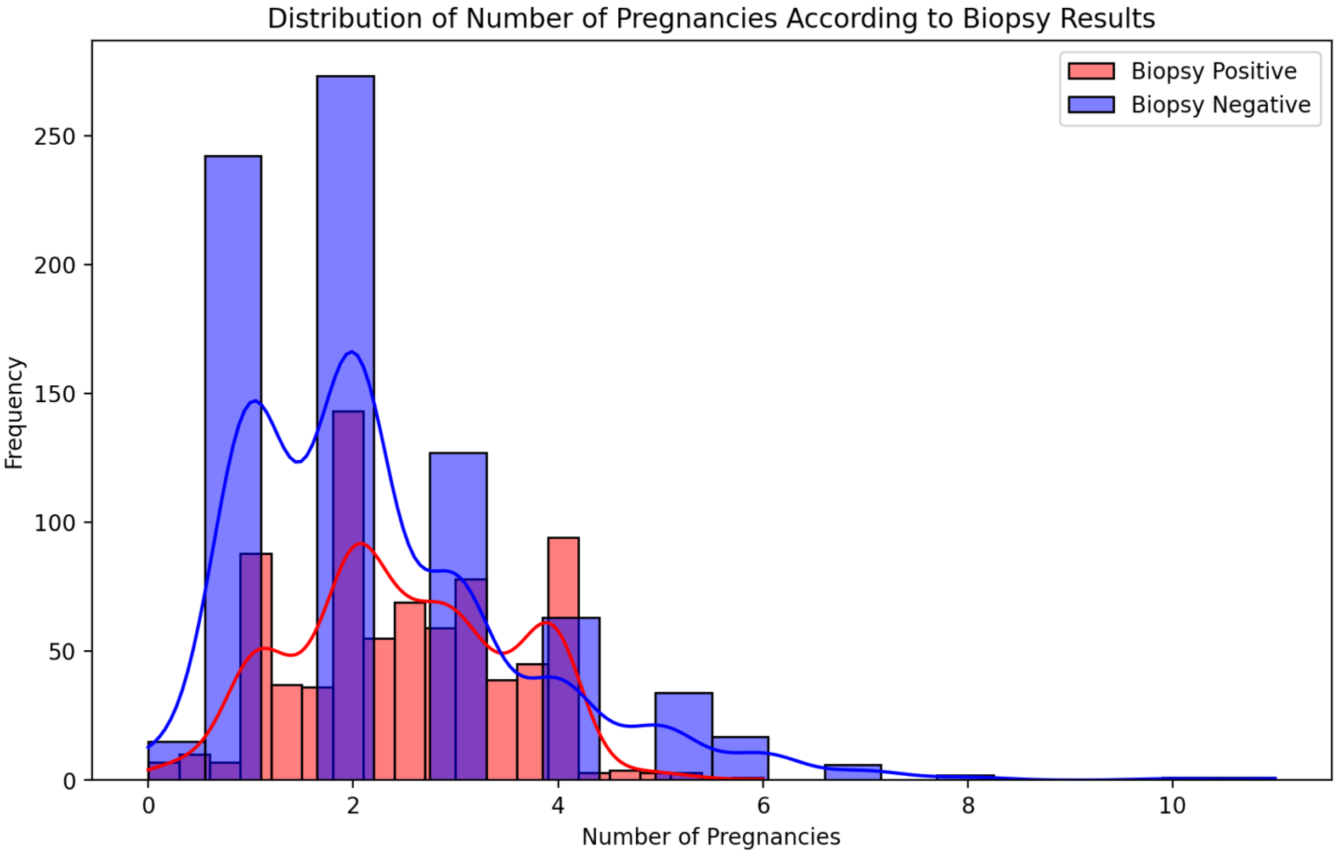
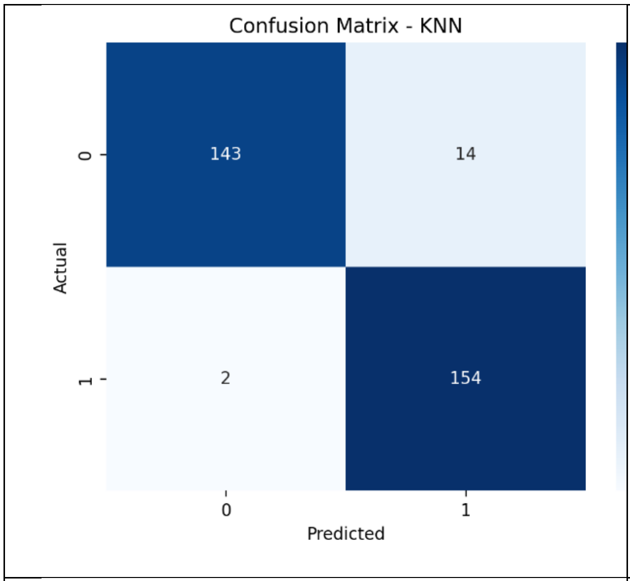
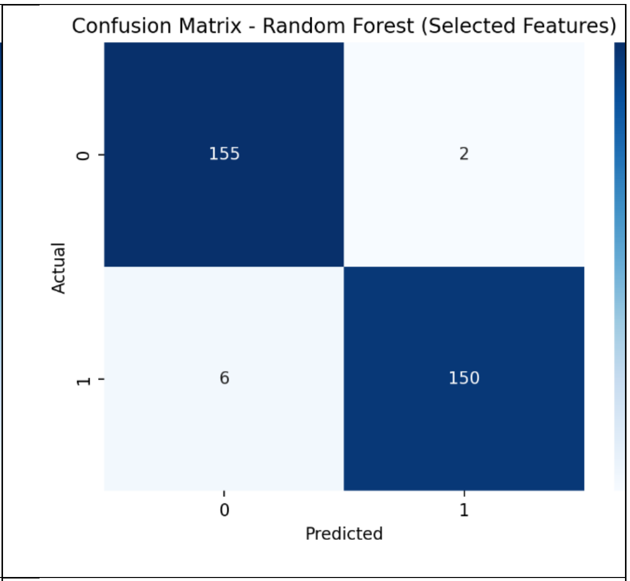


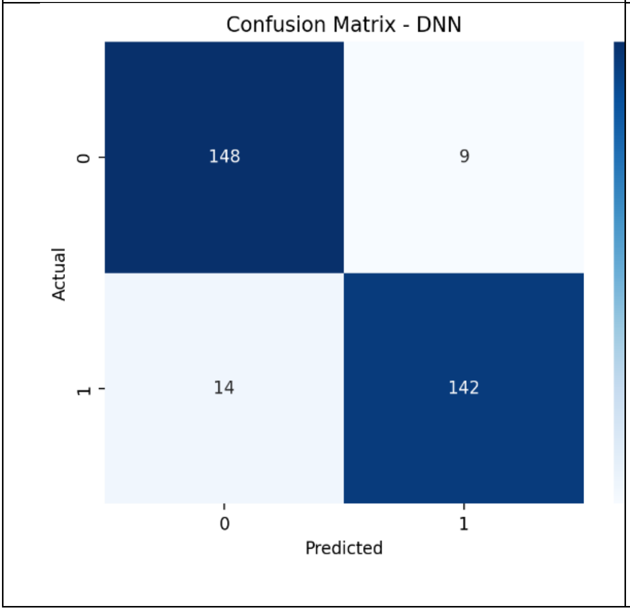
Figure 4: Distribution of number of Pregnancies According to Biopsy Results

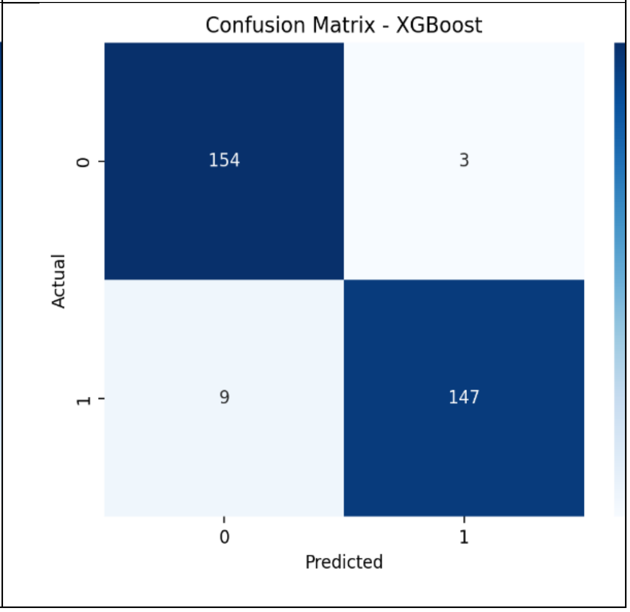
Figure 4 shows how the number of pregnancies were distributed in relation to the outcome of the biopsy, and the result of the biopsy was found to be positive or negative with respect to cervical cancer. Based on the visualization, it is observed that cases that are biopsy-negative are more significantly concentrated at the lower end of the range of pregnancies, especially between 1 and 3. This shows that most of the patients who are not cervical cancer patients have fewer pregnancies. Conversely, the biopsy-positive cases show a somewhat broader distribution with noticeable frequencies of higher pregnancy counts especially within the range of 2-4.

The KDE curve of positive cases is more widely spread, indicating that there is more variability among those affected. Also, this relatively low pregnancy counts (greater than 4) are not very common but is more pronounced in the positive group than the negative group. This tendency could reflect a possible correlation between the number of pregnancies and the risk of cervical cancer, yet it is not necessarily deterministic. The two distributions overlap, and the shift towards higher pregnancy counts in the positive cases is indicative that the given feature provides valuable predictive information. Nevertheless, it must be taken into account in combination with other clinical and behavioral variables to facilitate classification.


Confusion Matrix - KNN
Actual
0
1
143
14
2
154
0
1
Predicted


Confusion Matrix - Random Forest (Selected Features)
Actual
0
1
155
2
6
150
0
1
Predicted


Confusion Matrix - DNN
Actual
0
1
148
9
14
142
0
1
Predicted


Confusion Matrix - XGBoost
Actual
0
1
154
3
9
147
0
1
Predicted

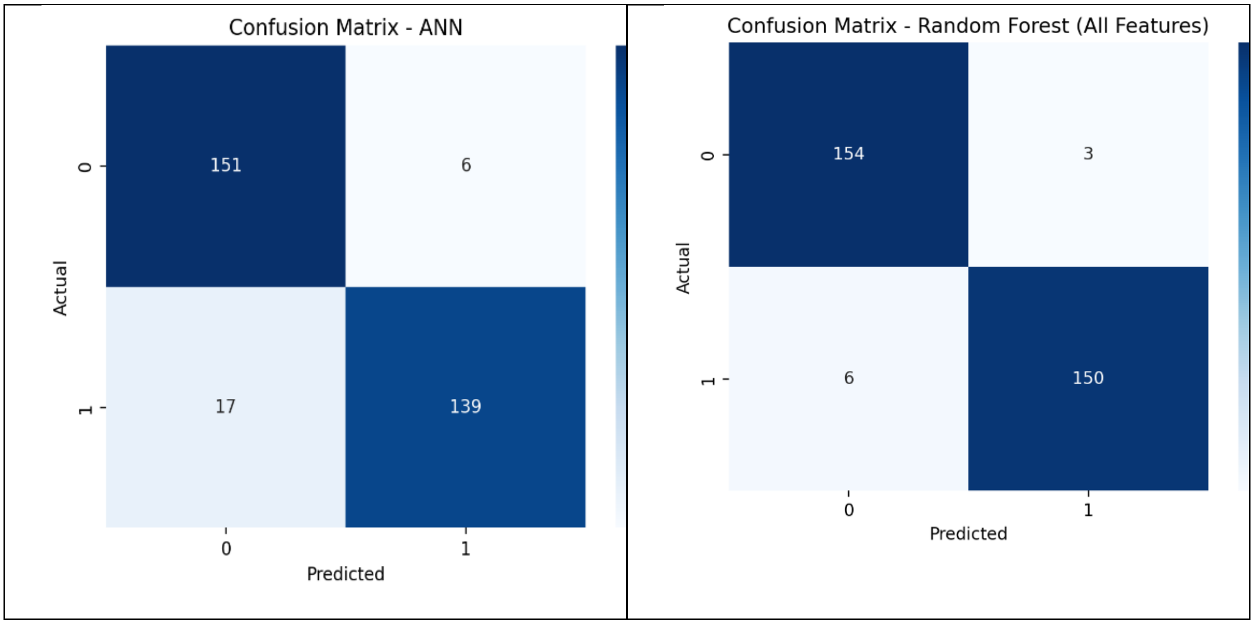


Figure 5: Confusion Matrix of Machine Learning Algorithms

The confusion matrices shown in Figure 5 can further explain how each of the models is performing beyond the issue of accuracy, especially concerning the types of errors that are most critical in medical diagnosis. Random Forest (with selected features) turns out to be the most reliable among all the models. It has the optimal balance of an accurate detection of positive and negative cases and the minimum false alarms and misdetections. It means that the noise has been successfully eliminated because feature selection and has enhanced the ability of the model to generalize. Random Forest (all features) and XGBoost models also show good results, being similar to the selected-feature Random Forest. Their ensemble character enables them to gather the sophisticated patterns in the data and they are very effective in structured medical data.

The Deep Neural Network (DNN) is a good performer but not better than the ensemble methods. Although it is able to capture nonlinear relationships, the rather small size of the data set has a minor impact on its performance, thus restricting its generalization capacities. Artificial Neural Network (ANN) demonstrates more negative results, especially with positive cases, which makes it less sensitive.

Conversely, KNN is a good detector of positive cases but shows a higher number of false positives hence less useful in real practice where false alarms are undesirable. The most efficient are ensemble methods, particularly, the Random Forest with feature selection because it is more robust and stable, as well as it is better balanced between various forms of classification errors.

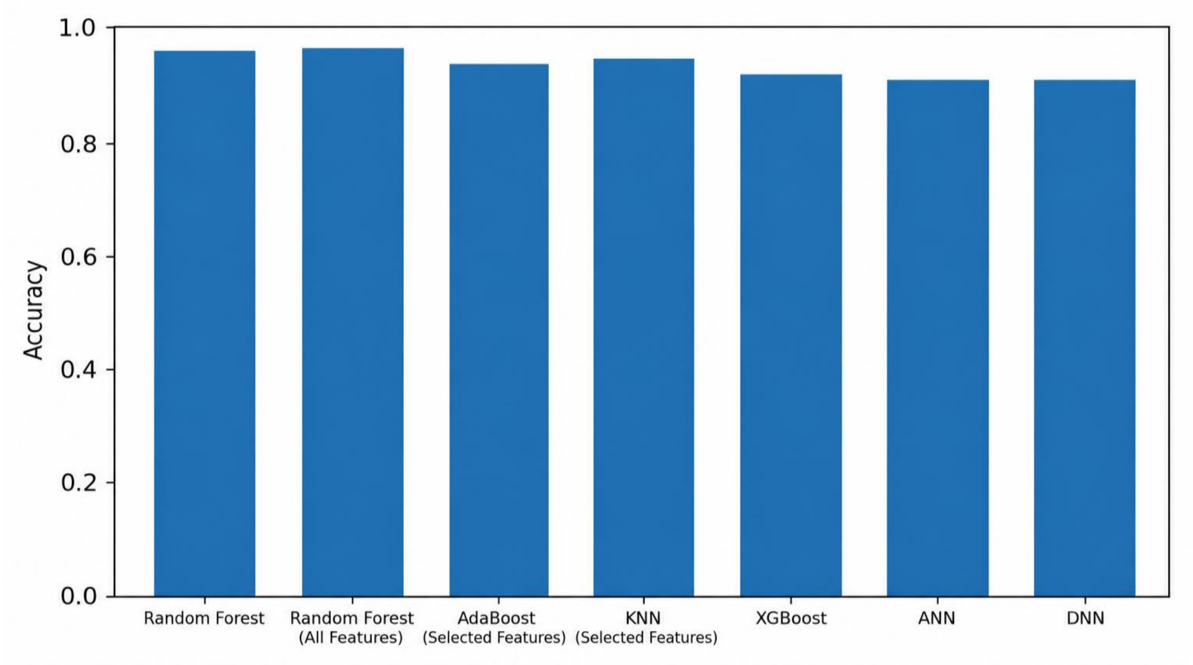


Figure 6: Accuracy Comparison with Different Algorithms

The bar chart shown in Figure 6 will be used to compare the classification accuracy of different machine learning and deep learning models in predicting cervical cancer. The accuracy of all models is high, which means that preprocessing is good and the data is well structured. Random Forest with the selected features among them outperforms the rest, proving that feature selection can optimize the efficiency of the model and remove irrelevant data, as well as improve generalization. Random Forest (all features) and XGBoost models are also very close to each other, which proves the effectiveness of the ensemble approach to structured medical data. Conversely, KNN, ANN and DNN have a little lower accuracy. Generally, the findings indicate that the best and most consistent and precise predictions in this task are obtained using ensemble methods, especially with the Random Forest and feature selection.

5. Conclusion

The current research compared several machine learning and deep learning systems in predicting cervical cancer based on various performance metrics in percentages. The results showed that the overall performance of Random Forest with selected features was the best with the accuracy of 97.44%, specificity of 98.73%, precision of 98.68%, F1-score of 97.40%, and lowest RMSE of 15.99, which shows that the features used enhanced the efficiency and minimized the prediction error. The accuracy and F1-score of the Random Forest (all features) model were also good, with an accuracy of 97.12% and F1-score of 97.09%, which proves the strength of the ensemble approach.

On the same note, XGBoost also competed with accuracy of 96.17%, a high specificity (98.09%) and a high precision (98.00%), which means that it gives out reliable predictions. KNN model had the best sensitivity and recall (98.72%), and it is very effective in identifying positive cases, but its specificity (91.08%) is lower, implying that there are more false positives. On the contrary, the ANN and DNN models demonstrated lower performance, at 92.33% and 92.65% accuracy, respectively, which implies that they are not as effective with smaller structured datasets. In general, ensemble models, especially Random Forest with feature selection are the most suitable in terms of accuracy, reliability and clinical applicability.

References

1. M. Arbyn, E. Weiderpass, L. Bruni, S. de Sanjose, M. Saraiya, J. Ferlay, and F. Bray, "Estimates of incidence and mortality of cervical cancer in 2018: A worldwide analysis," Lancet Global Health, vol. 8, no. 2, pp. 191-203, Feb. 2020.
2. H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA, A Cancer J. Clinicians, vol. 71, no. 3, pp. 209-249, May 2021.
3. F. J. Altaf and S. T. Mufti, "Pattern of cervical smear abnormalities using the revised Bethesda system in a tertiary care hospital in Western Saudi Arabia," Saudi. Med. J., vol. 33, no. 6, pp. 9-634, 2012.

4. J. Mondal and S. Joardar, "A comparative study to detect cervical dysplasia by using pap smear images," in Proc. 4th Int. Conf. Commun., Devices Comput., vol. 1046. Singapore: Springer, 2023, p. 113.
5. S. S. Gill et al., "Transformative effects of IoT, blockchain and artificial intelligence on cloud computing: Evolution, vision, trends and open challenges," Internet Things, vol. 8, Dec. 2019, Art. no. 100118.
6. J. Jeyshri and M. Kowsigan, "A comprehensive assessment of recent advances in cervical cancer detection for automated screening," in Image Processing and Intelligent Computing Systems. CA, USA: Taylor & Francis, 2023, pp. 171-184.
7. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436-444, 2015.
8. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Proc. Adv. Neural Inf. Process. Syst., vol. 25, 2012, pp. 1-9.
9. C. Szegedy et al., "Going deeper with convolutions," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2015, pp. 1-9.
10. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 770-778.
11. M. M. Gupta, "Forty-five years of fuzzy sets and fuzzy logic - A tribute to Professor Lotfi A. Zadeh," Scientia Iranica, vol. 18, no. 3, pp. 685-690, Jun. 2011.
12. N. Al Mudawi and A. Alazeb, "A model for predicting cervical cancer using machine learning algorithms," Sensors, vol. 22, no. 11, p. 4132, May 2022.
13. O. O. Umeonwuka, B. S. Adejumobi, and T. Shongwe, "Deep learning algorithms for RF energy harvesting cognitive IoT devices: Applications, challenges and opportunities," in Proc. Int. Conf. Electr., Comput. Energy Technol. (ICECET), Jul. 2022, pp. 1-6.
14. A. Gates et al., "Screening for the prevention and early detection of cervical cancer: Protocol for systematic reviews to inform Canadian recommendations," Systematic Rev., vol. 10, no. 1, pp. 1-22, Dec. 2021.
15. S. U. Rehman et al., "Personalised Comfort: A personalised thermal comfort model to predict thermal sensation votes for smart building residents," Enterprise Inf. Syst., vol. 16, no. 7, Jul. 2022, Art. no. 1852316.
16. J. Lu, E. Song, A. Ghoneim, and M. Alrashoud, "Machine learning for assisting cervical cancer diagnosis: An ensemble approach," Future Gener. Comput. Syst., vol. 106, pp. 199-205, May 2020.
17. A. Khamparia, D. Gupta, J. J. P. C. Rodrigues, and V. H. C. de Albuquerque, "DCAVN: Cervical cancer prediction and classification using deep convolutional and variational autoencoder network," Multimedia Tools Appl., vol. 80, no. 20, pp. 30399-30415, Aug. 2021.
18. M. Mehmood, M. Rizwan, M. G. Ml, and S. Abbas, "Machine learning assisted cervical cancer detection," Frontiers Public Health, vol. 9, Dec. 2021, Art. no. 788376.
19. M. Alsalatie et al., "A new weighted deep learning feature using particle swarm and ant lion optimization for cervical cancer diagnosis on pap smear images," Diagnostics, vol. 13, no. 17, p. 2762, Aug. 2023.
20. T. Aljrees, "Improving prediction of cervical cancer using KNN imputer and multi-model ensemble learning," PLoS ONE, vol. 19, no. 1, Jan. 2024, Art. no. e0295632.
21. K. M. M. Uddin et al., "An ensemble machine learning-based approach to predict cervical cancer using hybrid feature selection," Neurosci. Informat., vol. 4, no. 3, Sep. 2024, Art. no. 100169.
22. D. Ding et al., "Machine learning-based prediction of survival prognosis in cervical cancer," BMC Bioinf., vol. 22, no. 1, p. 331, Dec. 2021.
23. S. Uddin, I. Haque, H. Lu, M. A. Moni, and E. Gide, "Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction," Sci. Rep., vol. 12, no. 1, p. 6256, Apr. 2022.
24. R. L. Siegel, K. D. Miller, N. S. Wagle, and A. Jemal, "Cancer statistics, 2023," CA: A Cancer Journal for Clinicians, vol. 73, no. 1, pp. 17-48, 2023.
25. P. Bhardwaj, S. Kumar and Y. Kumar, "Deep Learning Techniques in Gastric Cancer Prediction and Diagnosis," 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON), Faridabad, India, 2022, pp. 843-850.
26. S. Chatterjee, D. Goyal, A. Prakash, and J. Sharma, "Exploring healthcare/health-product ecommerce satisfaction: A text mining and machine learning application," Journal of Business Research, vol. 131, pp. 815-825, 2021.
27. J. Lu, E. Song, A. Ghoneim, and M. Alrashoud, "Machine learning for assisting cervical cancer diagnosis: An ensemble approach," Future Generation Computer Systems, vol. 106, pp. 199-205, 2020.
28. 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, 2020.
29. I. J. Ratul et al., "Early risk prediction of cervical cancer: A machine learning approach," in 2022 19th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), pp. 1-4, IEEE, May 2022.