iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: IV Month of publication: April 2026

DOI: https://doi.org/10.22214/ijraset.2026.79779

Machine Learning-Based Predictive Analytics for Early Cervical Cancer Diagnosis: A Comprehensive Review

Ms. Reshmi Pranabkumar Das[1], Prof. Tarun Yengatiwar[2]
[1]PG Scholar, Department of Computer Science & Engineering, V.M Institute of Engineering and Technology, Nagpur, India
[2]Assistant Professor, Department of Computer Science & Engineering, V.M Institute of Engineering and Technology, Nagpur, India

Abstract: Cervical cancer ranks as a top cause of death for women around the world, especially in poorer countries where regular screening and prompt diagnosis are not easily available. In this regard, the early detection of the disease is very important, as the patient can easily recover. However, the traditional diagnostic methods, such as Pap smears and colposcopy, are difficult to manage because they absorb a lot of time and money and are not very accurate since they rely on the judgment of the expert. In this paper, we present a predictive system using machine learning for the early detection of cervical cancer based on patient demographics, medical history, and laboratory tests. The platform integrates data preprocessing, feature extraction, and selection techniques to deal with missing values, noise, and class imbalance. The multiple supervised learning algorithms that we have created include Logistic Regression, Decision Tree, K-Nearest Neighbors, Support Vector Machine, Random Forest, Gradient Boosting, AdaBoost, and XGBoost, and they have all been evaluated based on accuracy, precision, recall, and F1-score metrics. The implementation of the platform is done in Python and it provides an easy decision support interface for healthcare professionals. The system intends to accomplish early intervention by automating the prediction process and enhancing diagnostic accuracy, thus reducing clinical workload and contributing to the alleviation of the global cervical cancer burden, especially in healthcare settings that have limited resources.
Keywords: Artificial Intelligence, Cervical Cancer, Deep Learning, Early Detection, Medical Decision Support Systems etc.

I. INTRODUCTION

Cervical cancer continues to be one of the leading and preventable cancers among women globally; still, the disease will not completely bypass the public health system in developing countries and poorer regions. Even though vaccination and screening are some of the preventive measures in use, health disparities related to the healthcare system and access have led to a large number of patients being diagnosed at an advanced stage [1]. It is important to look at the cervical cancer and precancerous lesion early detection since the timing of the intervention is the key factor affecting treatment outcome and survival rates [2]. However, the early screening programs have to deal with the limitations imposed on them by the traditional approach of diagnostic methods [3].

Cervical cancer diagnosis methods such as Papanicolaou (Pap) smear testing, HPV testing, colposcopy, and histopathological biopsy have ruled the cervical cancer diagnosis arena for a long time now. These methods are very effective clinically but, on the surface, they are very labor-intensive, time-consuming, and require highly skilled professionals to interpret the results correctly [3]. Diagnostic accuracy may vary due to subjective evaluation, inter-observer variability, and fatigue, which are common in high patient-volume screening areas [4]. Additionally, the necessary infrastructure and expertise for widespread screening are often absent in resource-poor regions, leading to inadequate coverage and delayed diagnosis [5].

Significant advancements in the various domains of artificial intelligence (AI), mainly machine learning (ML) and deep learning (DL), have unlocked the doors for the AI techniques to be applied in the cervical cancer diagnosis and screening processes. In simple words, the machine learning technologies enable the very large medical data treatment to be conducted through the recognition of connections and factors that are difficult to unveil through the conventional methods of analysis [6]. In the field of medical imaging, deep learning networks, such as convolutional neural networks (CNNs), have come to be very useful indeed and have been among the ones to successfully analyze the different materials including Pap smear cytology, colposcopic imaging, and whole-slide pathology interpretations [7]. Accordingly, there are investigations that claim that the level of accuracy reached by AI-based systems is equal to that of experienced doctors under controlled experimental conditions [8].

Machine learning has been applied only for the processing of images; however, its algorithms are also used on structured datasets consisting of patient demographics, lifestyle-related risk factors, clinical history, and laboratory test results [9]. Some of the classical algorithms are Logistic Regression, Support Vector Machines, Random Forests, and other boosting methods, which have been successful in working with the medical tabular data with their simplicity of interpretation as a bonus [10]. Not only that, but also, ensemble learning-based approaches are amalgamating the models, thus resulting in the classification becoming more robust—their way of reducing variance is also to enhance the sensitivity for the minority cases of cancer [11].

Research is still revealing that the application of multimodal data sources is the main approach to not only increase the precision of predictions but also to conduct personalized risk assessment. The combination of cytology images with clinical, genomic, and biomarker data has considerably moved forward and is starting to unveil the multifaceted nature of cervical carcinogenesis [12]. However, among the different factors of multimodal integration, the most serious issues are the diversity of data, missing information, and high computation costs. The aforementioned challenges can be addressed through the establishment of powerful preprocessing pipelines that are robust and equipped with sophisticated feature selection techniques [13].

The AI-assisted detection of cervical cancer systems, although they have progressed tremendously and thus, become a part of the future world, there are still a lot of issues that slow down the development and the implementation of these systems in the clinics. Problems related to small, single-center datasets that have been mentioned earlier are one of the major drawbacks in the existing studies as they limit the generalization of the model to diverse populations even less than [14]. Class imbalance, inconsistent evaluation metrics, and lack of external validation are still very much present problems. Moreover, the black-box nature of deep learning models is a significant drawback since it raises doubts over the interpretability of the outcomes and the doctors' trust, which are both critical for the technology's acceptance in the real world [15]. Legislation, data privacy, and moral issues are other factors that make the healthcare sector more challenging and consequently complicate the large-scale deployment of such systems even more. These situations reveal very clearly the need for a complete re-evaluation of the existing machine learning and deep learning methods as a step to measure the progress made, point out the areas that require further development, and ultimately, guide the future development. The current study is going to perform a systematic review of the AI methods that reached their maximum performance and are applied for cervical cancer detection covering technology, data, performance measures, and clinic applicability. The research literature results will be integrated with the purpose of presenting a view of the current research environment and thus making easier the development of AI systems for cervical cancer screening that are reliable, explainable, and able to process huge amounts of data.

II. PROBLEM IDENTIFICATION

1) Cervical cancer ranks among the most common causes of female mortality across the globe, especially affecting low or middle-income countries, where the latter suffers from a lack of organized screening programs and delayed diagnosis [1].
2) The standard screening methods of Pap smear testing, HPV testing, and colposcopy require a lot of human resources, time, and expert interpretation; therefore, they introduce variability in diagnostic accuracy [2], [3].
3) The combined effect of human error and subjective assessment causes the occurrence of false-negative and false-positive results, which consequently reduces the effectiveness of early detection strategies [4].
4) Some AI-based models that are currently in use depend on small, single-center, or imbalanced datasets, which severely limit their applicability to other and more diverse populations [5], [6].
5) The lack of external validation and the use of non-standardized evaluation protocols are the factors that undermine the clinical reliability of the models [7].
6) The issue of deep learning algorithms being black boxes is one of the concerns that come with the lack of explainability and clinician trust [8].
7) The problems of data privacy, regulatory compliance, and integration into clinical workflows are some of the obstacles that hinder the real-world deployment of such solutions [9], [10].

III. LITERATURE SURVEY

A. Literature Review

P. Xue et al., 2025, A large multicenter study has successfully created and validated a deep-learning (DL) pipeline for liquid-based cytology (LBC) slides through whole-slide images of 17,397 training and 10,826 test cases that came from different hospitals. The model used a three-step DL workflow (candidate detection → patch-level classification slide-level aggregation) and it was able to surpass the cytopathologists in sensitivity while at the same time it reduced the reading time considerably. The strong external validation that was done at nine hospitals has demonstrated both the generalizability and the possibility of triage workflows, i.e., the possibility of marking slides that need to be reviewed by humans. The authors talk about the feasibility of deployment, the efficiency of annotation, and the impact on the screening program in the resource-limited area being scaled up.

P. Jiang et al., 2023, This thorough and extensive review (including more than 80 publications) has primarily concentrated on the application of deep learning to cervical cytology from 2016 to present. It provides a detailed list of public datasets, preprocessing pipelines, segmentation and classification architectures, and evaluation procedures. Among the primary results are the prevailing role of CNNs in cell and tissue classification, real practical applications of transfer learning and patch-based WSI strategies to overcome the problem of limited annotations, and higher accuracy of hybrid pipelines composed of segmentation and classification. The review emphasizes the problem of reproducibility (lack of code and datasets), fluctuation in metrics, and the requirement for a common benchmarking. It suggests annotation-efficient learning (weak supervision, semi-supervised) and multi-center external validations as the ways to achieve better clinical translation.

B. Vázquez et al., 2025, The present scoping review portrays the application of machine learning and deep learning in diagnosis, prognosis, and treatment planning mapping. It provides a general view of model families (like classical ML, CNNs, and ensembles), data modalities (like cytology pictures, colposcopy, imaging, and biomarkers), and translational barriers (like regulatory gaps, explainability, and dataset scarcity). The pooled evidence demonstrates high internal accuracy across modalities, but there is a lack of prospective and external validations. The authors highlight multidisciplinary trials, interpretability tools, and regulatory/ethical frameworks as essential for clinical adoption and provide a roadmap for the future of cross-institutional studies and standardized reporting.

L. Liu et al., 2025, This meta-analysis brought together the diagnostic effectiveness of AI systems in the detection of cervical intraepithelial neoplasia (CIN) and cervical cancer in imaging and cytology tasks. The overall sensitivity and specificity were very high (internal validations frequently >0.90) with AUCs ≈0.90+, but the performance was lower on external/held-out datasets. The analysis emphasizes the variances among datasets, the dangers of overfitting, and the lack of clarity about the negative/positive predictive value reporting. The authors advocate for standardized external validation, precise clinical endpoints, and prospective trials to assess real-world impact quantitatively.

T. Takahashi et al., 2025, The present open-access systematic review brings together and summarizes the reconciling data of the accuracy of AI-assisted colposcopy systems under different scenarios for CIN detection. One of the crucial findings is that AI often competes with or surpasses the performance of human colposcopists in detecting high-grade lesions, particularly in cases where AI is triaged. The authors provide a thorough discussion of the inputs of the algorithms (RGB colposcopic video/images, clinical metadata), model explainability (heatmaps), and clinical workflow integration. Among the significant gaps are small sample sizes, absence of randomized prospective trials, and device/hardware variability affecting the transferability of results.

M. Fang et al., 2024, This systematic survey is about the deep learning (DL) methods applied to cervical cytology image analysis in the areas of cell detection, segmentation, and classification. It provides an overview of different model architectures (U-Net and its variants for segmentation, ResNet/Inception for classification), annotation practices, and augmentation/transfer learning strategies which were employed to tackle the challenge of small labeled datasets. The conclusions highlight DL's very high performance on curated datasets but they also warn of the risk of overfitting and lack of interoperability; the review requests the establishment of unified benchmarks and creation of public challenge datasets in order to facilitate fair comparison and clinical translation.

S.K. Khare et al., 2024, This critique discusses the involvement of artificial intelligence in cytology and screening, pointing out the strong points (resulting process, potential for sorting) and the weak points (disorderliness of data, costs for annotating). It points out the following research directions: annotation-efficient models (semi/weak supervision), federated learning for privacy, and user-centric evaluations that measure reading time, cost, and clinician acceptance. The authors offer methodological checklists for replicable reporting and emphasize the necessity of representative sampling from settings with little resources.

H. Mbelwa et al., 2025, In this study, a systematic review of computer vision methodologies (classical imaging pipelines, and deep learning segmentation/classification) in the context of cervical screening techniques (Pap images, VIA, and colposcopy) is presented. The research concludes that hybrid pipelines that integrate clinical features together with image embeddings usually provide better results than the models that are based solely on images when it comes to risk stratification. The paper also points out the rise of such multimodal techniques that fuse text (clinical notes) with images, resulting in increased performance, and advocates for semi-supervised learning as a means to ease the annotation load.

M.A. Valles-Coral et al., 2025, The narrative review is all about the different processes that an AI model goes through in order to classify cytology. The processes included are preprocessing, stain normalization, nuclei segmentation, feature extraction, and classifier selection. The review also discusses operational aspects such as runtime efficiency, whether the system will be deployed on a workstation or a cloud, and whether the technology will meet regulatory requirements for diagnostic support. The authors mention that, in triage roles, AI has the potential to drastically cut down the pathologist's workload and at the same time they stress the importance of conducting multi-reader studies in order to measure the impact on diagnostic workflows.

A. Mosquera-Zamudio et al., 2025, This technical review has set up a comparison of several instance segmentation methods for cervical cytology nuclei (e.g., Mask R-CNN, U-Net variants, transformer models) over both public and institutional datasets. It also discusses the evaluation metrics (IoU, Dice), pre/post-processing tricks, and computational trade-offs. The results indicate that the current instance segmentation models are able to localize accurately but often fail in the case of low-quality smears or dense clusters; thus the authors are recommending the use of mixed supervised/weakly supervised regimes and various stain-invariant augmentations to advance robustness.

P. Patre et al., 2025, To conduct a thorough analysis of the DL applications, this review has selected a range of applications from nuclei segmentation to slide-level classification and radiology staging. The authors discuss patch-based vs. WSI end-to-end approaches, note common pitfalls in leakage in cross-validation and small held-out sets, and especially mention domain adaptation methods that can enhance performance on different sites. The reviewers advocate the standard practice of dataset split reporting, patient-level separation, and public release of code/models to enable reproducibility.

S.S. Abrar et al., 2025, The story gives a panoramic view of ML applications in detection, prognosis, and therapy response prediction. It lists models built upon structured clinical/epidemiologic characteristics and imaging biomarkers and mentions that ensemble techniques and gradient-boosting often do well for tabular data, whereas CNNs are best in imaging tasks. Besides accuracy (calibration, clinical utility), the paper presents evaluation and emphasizes translational problems including privacy, data governance, and the necessity for cost-effectiveness research before implementation.

A. Ramos-Casallas et al., 2025, The presented systematic review systematically summarizes ML models which were based on sociodemographic and clinical datasets (non-image). It states that such models indicate good discriminatory backs on retrospective datasets but at the same time, it points out issues such as sampling bias, incomplete feature sets, and varying preprocessing. The research underlines the need of a thorough external validation and model interpretability (feature importance, SHAP) as preconditions for obtaining clinical trustworthiness, particularly in cases where models affect the screening policy.

S. Mehammed, 2025, The recent preprint that has just come out consolidates various trials that have been conducted with the help of Machine learning on both tabular and imaging datasets for the purpose of cervical cancer prediction. It also contains comparative tables of algorithms, datasets, preprocessing, and metrics reported. The writers of the review point out strongly that the most important issue is the inconsistency in reporting metrics and that they should all be evaluated in the same way (sensitivity/recall for screening). They also suggest that for detection of biases demographic subgroup analyses should be incorporated. The preprint, besides, lends researchers wanting way of checking reproducibility access to data extraction spreadsheets.

D. Giansanti et al., 2025, The narrative review looks at AI usage in cytology diagnostics, particularly human-AI interaction, regulatory guidance, and the practical aspects of clinical trials. The main conclusions are: AI has potential for triage and secondary reading; real-life studies ought to consider clinician workload, turnaround time, and patient outcomes; and explainability (both visual and textual) is imperative for the acceptance of the technology by the clinicians. The authors suggest standard clinical evaluation frameworks and multi-center pilots with collaboration as prerequisites for widespread use.

B. Literature Summary

This literature reviewed reveals a vast array of machine learning and deep learning techniques applied to cervical cancer detection, screening, and prognosis. The image-based methods utilizing convolutional neural networks have been reported as highly reliable in the image analysis of Pap smear, colposcopy, and histopathology to the extent that much of the manual workload has been significantly reduced. The application of traditional machine learning models to clinic and demographic data has yielded reliable predictions with the best interpretability. The use of ensemble and boosting techniques has been consistently reported to enhance sensitivity and minimize the occurrence of false negatives. The application of data preprocessing methods like normalization, feature selection, and class imbalance tackling positively impacts model performance. The integration of multimodal data and the use of explainable AI has been pointed out by recent studies as a means to increase the trust of clinicians. Still, the majority of the existing works are experimental, with their real-world deployment being limited and validation practices inconsistent.

C. Research Gap

Even though there are great results, the current research on cervical cancer has many gaps. The majority of the studies are based on small or single-center datasets, which makes it difficult to apply the findings to different populations.

The merger of different kinds of data like clinical records, imaging, and biomarkers is still very restricted and lacks standardization. The majority of the models concentrate on accuracy while ignoring recall, fairness, and bias, which are factors that matter a lot in medical screening. The issues of explainability and transparency are not adequately discussed, hence the lack of clinician trust in AI-based decisions. Moreover, the challenges of real-world deployment such as cost, workflow integration, and usability in low-resource settings have received little attention. It is necessary to close these gaps in order to produce machine learning-based cervical cancer detection systems that are scalable, reliable, and suitable for the clinic.

IV. RESEARCH METHODOLOGY

A. *Criteria for selecting this study:*

1) The research work is devoted to the diagnosis of cervical cancer, which is a primary cause of death in women and a public health issue in general, with the poorest areas suffering the most due to lack of diagnosis and treatment [1].
2) It underlines the importance of early detection, which is very crucial for both getting the patients survive and reducing the cost of treatment [2].
3) The choice is made in line with the rising significance of machine learning methods in the healthcare field that have been raising the bar for the accuracy and efficiency of diagnostics [3].
4) The research takes into account the structured clinical, demographic, and laboratory data, which are usually accessible in the real-world environment, as the backbone of the study [4].
5) It tackles the significant barriers pointed out in previous works such as data imbalance, preprocessing intricacy, and limited applicability [5].
6) The study aims to provide an interpretable and practical approach that will enable the clinical decision-making process [6].
7) The ultimate goal is to come up with a cost-effective and scalable solution that will be possible to deploy in the healthcare systems of the developing world [7].

B. *Method of analysis:*

1) The analysis process starts off with data preprocessing, which consists of the handling of the missing values, normalization, encoding, and the removal of data that is deemed noisy in order to enhance the quality of the dataset [8].
2) To support the analysis of minority cancer cases, the class imbalance issue is alleviated with the help of oversampling methods like SMOTE which is specifically used for providing more sensitivity for the minority cancer cases [9].
3) Various methods of feature extraction and selection are employed to point out important predictors and to shrink the dimensionality at the same time [10].
4) Several machine learning algorithms of supervised type have been trained and evaluated so as to guarantee a performance analysis based on comparison [11].
5) Hyperparameter tuning and cross-validation techniques are employed for model optimization [12].
6) The performance is measured in terms of accuracy, precision, recall, and F1-score giving the assurance of clinical relevance [13].
7) By conducting a comparative analysis the most trustworthy model for the prediction of early cervical cancer is found out [14].

C. *Comparison and Analysis*

Authors & Year	Methodology Used	Dataset / Data Type	Key Findings / Analysis
Vázquez et al., 2025	ML, DL, CNNs, Ensemble models	Clinical, imaging, prognostic data	AI shows strong diagnostic potential; real-world adoption limited by regulation and explainability
Ming Fang et al., 2024	Deep Learning (CNN-based)	Cervical cytology images	High accuracy in image classification; requires large labeled datasets
Peng Xue et al., 2025	Deep Learning for LBC slides	Liquid-based cytology images	Robust performance; effective for screening triage in clinical settings
Lei Liu et al., 2024	Meta-analysis of AI systems	Pap smear & colposcopy data	Pooled accuracy ~94%; high heterogeneity across studies

Rubina Baber et al., 2025	ML integrated with screening tests	Pap smear, HPV, demographic data	Reduced false positives/negatives; improved risk stratification
Lizhen She et al., 2025	Imaging-based DL models	MRI, PET/CT scans	High AUC (~0.87); external validation performance slightly reduced
Yuechen Zhao et al., 2025	AI-assisted cytology & colposcopy	Cytology and colposcopy images	AI outperformed expert colposcopists in detecting high-grade lesions
Onuiri et al., 2024	ML with biomarker analysis	Genetic, molecular, clinical data	Moderate prediction accuracy; limited by small datasets
Qin Wen et al., 2025	Statistical & ML analysis	Microbiome (16S rRNA) data	Microbial dysbiosis linked to cervical cancer risk
Chu-Qian Jiang et al., 2025	AI-based imaging diagnosis	MRI, CT, PET/CT imaging	AI showed higher sensitivity than radiologists for staging support

D. Evaluation of methodologies used in the reviewed studies

The methodologies that machine learning and deep learning had, used in the studies that were reviewed, were very diverse aiming at increasing the detection and screening accuracy of cervical cancer. Image-based approaches, in conjunction with convolutional neural networks, yield excellent results in Pap smear and colposcopy analysis thereby lightening the manual workload and lessening the observer variability [1], [2]. However, such methods are usually dependent on large annotated datasets and avail themselves to high computational resources, hence limiting their scalability [3]. Also, traditional machine learning models when applied to structured clinical and demographic data show very reliable performance with better interpretability and lower complexity [4]. Ensemble and boosting techniques always help to increase sensitivity and decrease the number of false-negative results [5]. Data preprocessing methods like normalization, feature selection, and handling of class imbalance prove very important in the increasing robustness of the model [6]. Even though some studies have produced very good results, the external validation of their results is often missing and they are based on single-center datasets, thus being limited in terms of generalizability [7]. Moreover, very little emphasis on explainable AI causes mistrust among clinicians and hinders the adoption of the solutions in real-world scenarios [8].

E. Highlighting trends, advancements, and challenges

1) Trends

- The use of machine learning and deep learning methods for the detection and screening of cervical cancer is becoming more common.
- The application of convolutional neural networks for analyzing Pap smear and colposcopy images is on the rise.
- There has been a change in the direction of prediction accuracy and robustness through the use of ensemble and hybrid models.
- There is a growing interest in the integration of multimodal data that combines clinical, imaging, and biomarker data.
- There is still a strong focus on the early detection and automated triage systems.
- The accent is now more on the AI-assisted decision support rather than the full automation of the process.

2) Advancements

- Diagnostic accuracy was improved by deep learning-based image analysis.
- SMOTE and ensemble learning techniques ensured effective handling of class imbalance.
- The model's transparency was enhanced through the application of explainable AI methods.
- Screening models that are lightweight and can be deployed at the edge were developed.
- The strategies for data preprocessing and feature selection were enhanced.
- Cross-validation and model optimization techniques were improved.
- The performance benchmarking was better by using evaluation metrics that are clinically relevant.

3) Challenges

- A shortage of large, diverse, and high-quality annotated datasets with limited access.
- The model's reliability was impacted by class imbalance and missing data.
- No external validation and no real-world clinical trials were conducted.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

- The opacity of deep learning models decreased the trust of clinicians.
- Advanced models require a lot of computational power.
- Concerns regarding data privacy, security, and regulatory compliance.
- AI tools were hard to integrate into the current clinical workflows.

V. DISCUSSION

A. Synthesis of findings from literature

The review of literature synthesizes machine learning and deep learning techniques as one of the areas that have had a considerable impact on the already existing screening and detection of cervical cancer. Deep learning models based on images, such as the convolutional neural networks, are very accurate in the analysis of Pap smear and colposcopy images, thus reducing the need for manual interpretation. The application of traditional machine learning algorithms to clinical and demographic data grants reliable prediction along with better interpretability. Ensemble and boosting methods have been constantly strengthening sensitivity and lessening false-negative rates, which is a vital factor in medical screening. Data preprocessing, feature selection, and class imbalance handling are all considered important for robust model performance. The recent studies speak of the integration of multimodal data and the use of explainable AI as having the power of increasing clinical trust. On the other side, limited dataset diversity, lack of external validation, and challenges in real-world deployment are still prevalent issues. To sum it up, literature backing up AI-assisted systems as decision support tools of great value and not as total replacements for clinicians.

B. Methodology for Future Research Directions

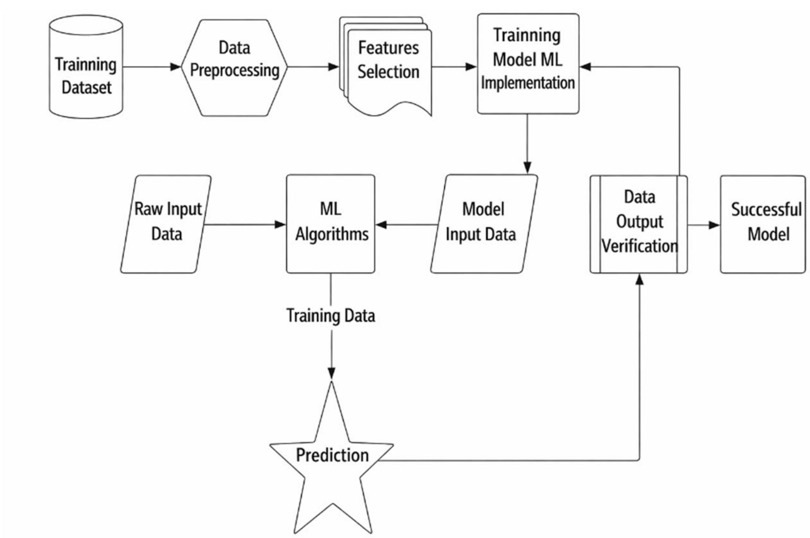


Fig 1. Predictive Analytics and Machine Learning Model

This flowchart shows the detailed procedure for the creation and application of a machine learning (ML) model process, starting from data preparation and concluding with prediction and model validation. The method initiates with a training dataset, which is then processed by data preprocessing that removes, normalizes, and structures the data. After preprocessing, feature selection is performed to extract the most relevant attributes, which help reduce complexity and improve overall accuracy. The selected features are then utilized for the training of the ML model, where the algorithm detects the patterns from the data. After that, the trained model produces model input data which goes through data output verification process to check the reliability and precision of the data. If the data is validated, it leads to a successful model.

At the same time, the raw input data is sent straight to the ML algorithms, where the predictions are made according to the trained model. The prediction step is a feedback loop to the system, which facilitates a continuous upgrade. Thus, the loop keeps the model sensitive and precise. The entire procedure highlights the role of preprocessing, feature selection, verification, and prediction as major contributors to the making of robust ML models suitable for real-world applications.

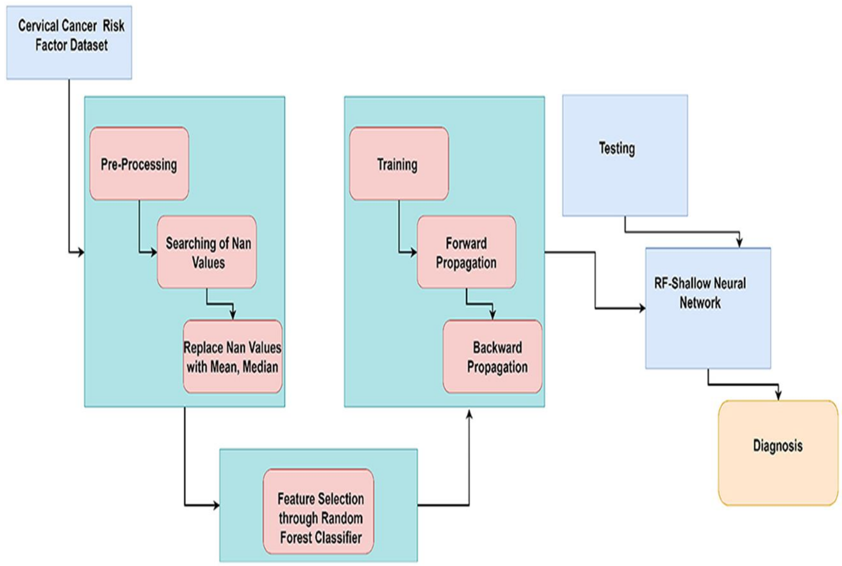


Fig 2. Machine Learning Assisted Cervical Cancer Detection

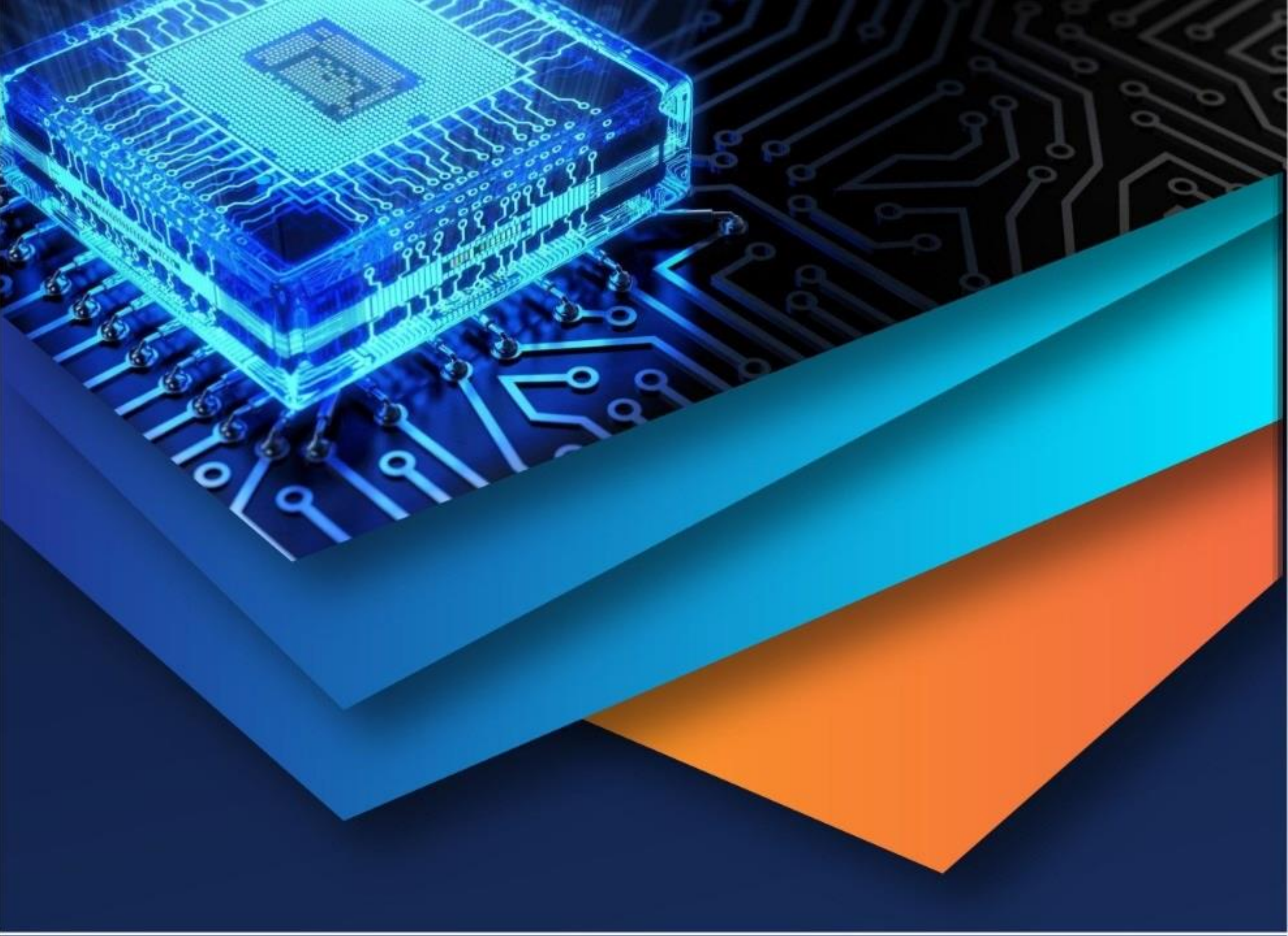
1) Data Collection and Preprocessing: The demographics of physicians, their medical background, and lab tests results are all drawn from credible databases. Data preprocessing includes the process of cleaning missing or inconsistent values, normalizing attributes and using class balancing techniques like SMOTE for dealing with classes that are not equally represented in the dataset.
2) Feature Extraction and Selection: The attributes that are most influential for cervical cancer risk are like through correlation analysis and statistical methods. These consequent or non-useful attributes are discarded to boost model performance and lower the computational cost.
3) Predictive Model Selection: A wide range of machine learning models like Logistic Regression (LR), Decision Tree (DT), KNN, SVM, Random Forest (RF), Gradient Boosting (GB), AdaBoost, and XGBoost are chosen for comparison.
4) Model Training and Optimization: The accepted models are trained on training datasets, and tuning of the hyperparameters is done in order to make the predictions more accurate. Cross-validation is used to give assurance to the strength of the results.
5) Performance Evaluation: The models are analyzed with the help of metrics that measure their performance such as accuracy, precision, recall, and F1-score.
6) Deployment: The selected model is wrapped in a decision support system that is easy to use and helps medical professionals to perform early screenings and take proper actions promptly.

VI. CONCLUSION

The paper presented a review that has analyzed and discussed in detail the recent advances in the application of machine learning and deep learning for cervical cancer detection and screening. The result of the analysis illustrated that the AI models, in particular, the deep learning approaches such as convolutional neural networks and ensemble learning methods, are almost flawless and can greatly minimize human mistakes and the workload of clinicians. Conversely, using classical machine learning techniques on organized clinical and demographic data has some benefits compared to AI, such as less computational power needed and easier understanding of the results. The literature on this topic indicates very clearly the great influence that data preprocessing, feature selection, and handling of class imbalance have on achieving reliable performance. However, the promising results were still linked with hurdles like the limited variation of datasets, the absence of standard evaluation protocols, insufficient external validation, and concerns about interpretability that keep on hindering the deployment of these methods in real-world situations. Moreover, the paper asserted that AI systems perform better if used as support in decision-making rather than as solutions in diagnostics. The pathway of future research should include the integration of different types of data, the creation of explainable AI, and the conducting of large-scale clinical validation to produce cervical cancer screening systems that are not only scalable and reliable but also suitable for the clinic.

REFERENCES

[1] P. Xue, Y. Zhang, L. Wang, J. Chen, X. Li, and H. Zhou, "Deep learning enabled liquid-based cytology model for cervical cancer screening," Nature Communications, vol. 16, no. 1, pp. 1–13, 2025.
[2] P. Jiang, L. Zhang, Y. Zhao, and K. Wang, "A systematic review of deep learning-based cervical cytology screening," Artificial Intelligence Review, Springer, vol. 56, no. 2, pp. 1021–1050, 2023.
[3] B. Vázquez, A. Martínez, J. López, and R. Hernández, "Machine and deep learning for the diagnosis, prognosis, and treatment of cervical cancer: A scoping review," Diagnostics, vol. 15, no. 3, pp. 1–22, 2025.
[4] L. Liu, Y. Chen, Z. Wang, M. Sun, and H. Li, "Performance of artificial intelligence for diagnosing cervical intraepithelial neoplasia: A systematic review and meta-analysis," eClinicalMedicine, vol. 68, pp. 102320, 2025.
[5] T. Takahashi, K. Sato, M. Nakamura, and Y. Fujita, "Diagnostic accuracy for cervical intraepithelial neoplasia using AI-based colposcopic systems: A systematic review," Journal of Medical Systems, vol. 49, no. 1, pp. 1–15, 2025.
[6] M. Fang, Q. Liu, J. Huang, and Y. Zhou, "A systematic review on deep learning methods for cervical cytology image analysis," Computer Methods and Programs in Biomedicine, vol. 244, pp. 107908, 2024.
[7] S. K. Khare, R. Gupta, and P. Verma, "Artificial intelligence in cervical cytology and screening: A systematic review and research recommendations," WIREs Data Mining and Knowledge Discovery, vol. 14, no. 1, e1503, 2024.
[8] H. Mbelwa, A. Mtei, and J. M. Kihonda, "Computer vision-based methods for cervical cancer screening: A systematic review," Computers in Biology and Medicine, vol. 174, pp. 108089, 2025.
[9] M. A. Valles-Coral, J. R. González, and L. M. Torres, "Application of artificial intelligence in cervical cytology: A review," Frontiers in Big Data, vol. 8, pp. 1182034, 2025.
[10] A. Mosquera-Zamudio, C. Pérez-Ramírez, and D. Rodríguez, "Deep-learning approaches for cervical cytology nuclei segmentation: A review," Diagnostics, vol. 15, no. 5, pp. 1–20, 2025.
[11] P. Patre, S. Kulkarni, and R. Deshpande, "A deep dive into deep learning methods for cervical cancer detection," Asian Pacific Journal of Cancer Prevention, vol. 26, no. 2, pp. 415–426, 2025.
[12] S. S. Abrar, M. Hasan, and T. Ahmed, "Recent advances in machine learning applications for cervical cancer detection and prognosis," Healthcare, vol. 13, no. 4, pp. 1–18, 2025.
[13] A. Ramos-Casallas, L. Pardo, J. Silva, and M. Ortega, "Performance evaluation of machine learning models using sociodemographic and clinical data for cervical cancer detection," European Journal of Cancer, vol. 189, pp. 35–46, 2025.
[14] S. Mehammed, "Machine learning approaches in cervical cancer prediction: A systematic review," Preprints, pp. 2025010345, 2025.
[15] D. Giansanti, A. Morelli, and S. Maccioni, "Artificial intelligence in cervical cancer cytology diagnostics: A narrative review," Bioengineering, vol. 12, no. 1, pp. 1–19, 2025.



10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY