Machine Learning in Early Detection and Diagnosis of Cervical Cancer: Advancements and Clinical Implications

Saranya. A1, S. Ravi1*, Harsha Latha. P2, T. Kalaichelvi2
1,2 Department of Computer Science, School of Engineering and Technology,
Pondicherry University, Puducherry, 605014, India

Abstract. Cervical cancer ranks as the fourth most common cancer among women, accounting for approximately 7.9% of all female malignancies. A significant contributing factor has been the lack of proper early treatment, highlighting the critical need for prevention through early diagnosis. Cervical cancer remains one of the leading causes of mortality among women diagnosed with various forms of cancer globally. Early detection and accurate diagnosis greatly enhance survival rates among cervical cancer patients. Traditionally, diagnosis has relied on imaging tools such as CT and MRI scans. However, manual interpretation of these medical images is time-consuming and prone to inter-observer variability. Machine learning (ML) approaches have shown promising results in automating cervical cancer diagnosis in recent years. ML has emerged as a powerful tool across numerous medical applications, with classifiers effectively identifying diseases based on extracted features, such as detecting cancer cells in the cervix for early diagnosis. In this research, we explore ML-based approaches for cervical cancer prediction, examining the significance of factors like age, sexual activity, HPV infection, and smoking habits in predicting invasive cancers. This study provides a comprehensive overview of ML methods for cervical cancer diagnosis. In conclusion, using machine learning algorithms for detecting and diagnosing cervical cancer from medical images has yielded promising, automated results.

Keywords: Cervical cancer detection; Classification; Machine learning.

1. INTRODUCTION

Cervical cancer remains a significant global health concern, primarily affecting the cells of the cervix and most commonly caused by persistent infection with human papillomavirus (HPV). Despite advances in prevention through vaccination and regular screening methods such as Pap tests, cervical cancer continues to impact thousands of individuals each year, particularly women between the ages of 35 and 44. Early stages often present no symptoms, making preventive measures and early detection critically important. Treatment options, including surgery, radiation therapy, chemotherapy, and emerging targeted therapies, have substantially improved patient outcomes, especially when the disease is identified early. Continued research into the causes, risk factors, diagnostic methods, and treatment innovations is essential to reduce the global burden of cervical cancer further and improve the survival rate. The fatality rate and incidence of cervical cancer have fallen by 90% due to the Pap test. This test's shortcomings include a lack of a clinical component, poor patient compliance, inconsistent diagnosis, inefficient follow-up care, and carelessness by the specialists administering it because it is so time-consuming [1].

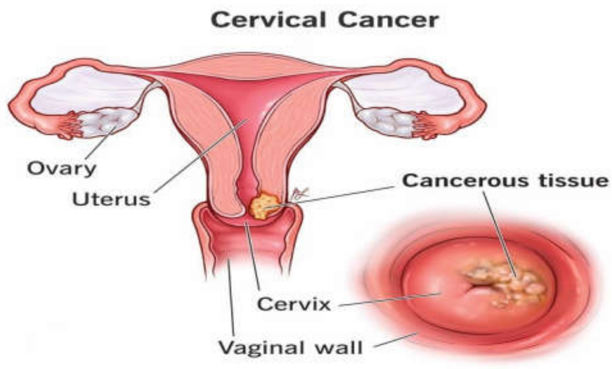

Fig 1. Diagrammatic Representation of Cervical Cancer

By introducing the pap smear test, the number of deaths from cervix cancer has decreased since the Human papillomavirus (HPV) vaccination helps to avoid infections below 18 years and in younger women. However, day by day, the mortality rate among developing countries is increasing rapidly. Fig 1. Explains cancers in women. A woman's cervix is where cervical cancer forms. The cervix is the constrictive part of a woman's uterus or womb. The cancer-causing human papillomavirus, or HPV, spreads through sexual contact. Sometimes, women are affected by HPV infection at some point; it goes away and becomes inoffensive.

Machine Learning came into the picture after many attempts, especially for the classification task, which helps to identify the detection of cervical cancer in earlier stages. Some widely used ML algorithms are divided into support vector machines, k-nearest neighbors, Random Forest trees (RFT), classification and regression trees (CART), and Artificial Neural Networks. The ML method has recently been used to diagnose medical imaging or cervical images using Computer-aided Diagnostic systems. Figure 2 explains Cervical Cancer Compared to other cancers in women. In case of earlier detection and precise diagnosis of cervical cancer, these systems can analyze medical imaging, patient data, and other clinical characteristics. While also lowering the cost and time associated with screening and diagnosis, implementing ML algorithms in CAD systems can increase the accuracy of cervical cancer diagnosis. Extensive patient data, clinical parameters, and ML datasets train the medical image data to find patterns and make precise predictions. Additionally, to recognize minute variations in the appearance of cervical tissue and find early-stage lesions that conventional methods would overlook [2].

The following section provides an overview of the most recent findings in using ML-based Computer Aided Design systems to identify cervical cancer. Recent studies have used machine learning (ML) techniques for cervical cancer diagnosis and the benefits and drawbacks of this approach. The potential areas for further research and ways to enhance the precision and effectiveness of ML-based Computer Aided Design systems for cervical cancer diagnosis are also pinpointed.

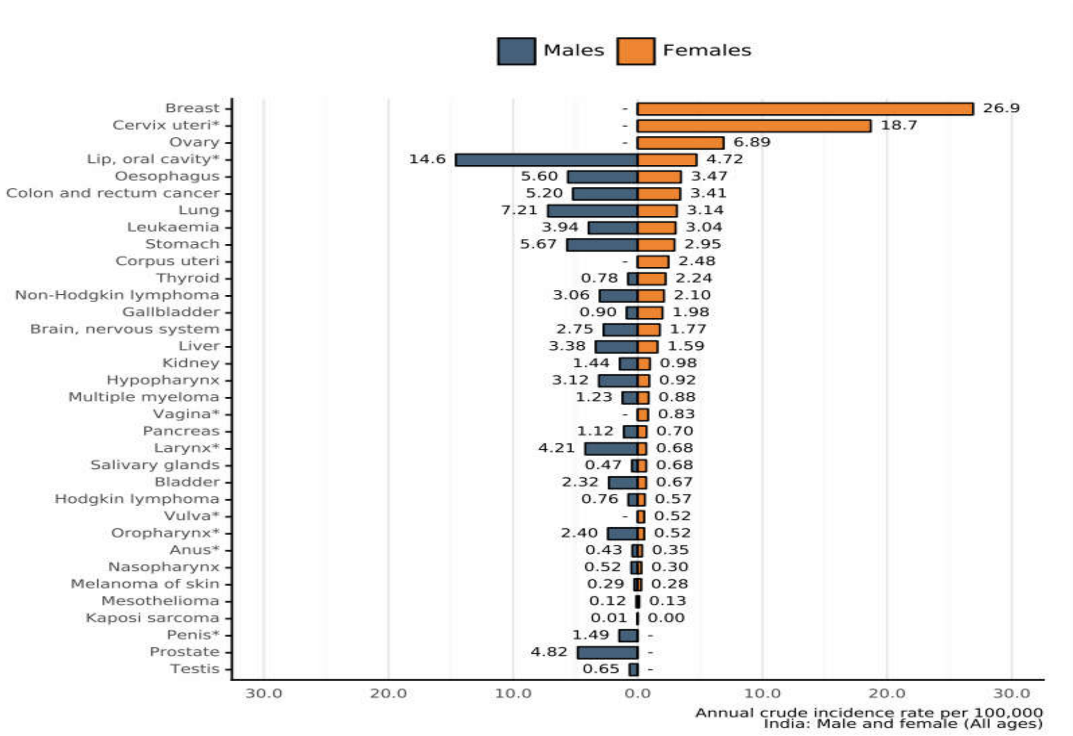

Fig 2. Cervical cancer compared to other cancers in women [3]

1.1 Overview of Medical Image Analysis and Medical Image Modalities

Medical image analysis obtains relevant information from medical photographs, frequently accomplished through computer algorithms. This analysis can be obtained using X-ray (2D and 3D), ultrasound, computerized tomography (CT), magnetic resonance imaging (MRI), nuclear imaging (PET and SPECT), and the microscope.

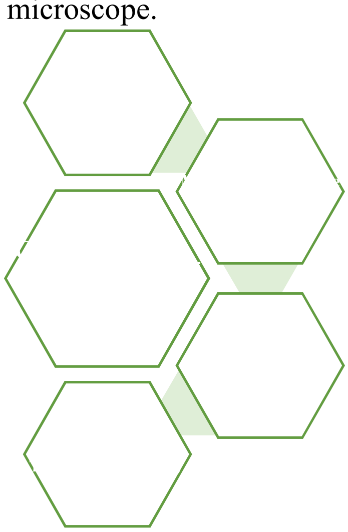

Fig. 3. Medical Imaging Modalities of Cervical Cancer

Medical segmentation of images is challenging because of the many constraints imposed by the medical image collection technique, pathology type, and biological variances. Figure 3 explains that Professionals can analyze medical images, and there is a need for more medical imaging professionals [4].

Several medical imaging modalities can be used in the detection and management of cervical cancer:

Ultrasound: First imaging modality to evaluate cervical cancer. Transvaginal ultrasound provides complete images of the cervix, uterus, and surrounding structures. When comparing imaging cervical Cancer, ultrasound beats out methods like CT or MRI in terms of cost, speed, and accessibility. The clinical use of ultrasonography for cervical cancer has been enhanced by new techniques like three-dimensional (3D) ultrasound and color Doppler. Fig 4 Explains the 2-dimensional grayscale ultrasound image of cervical carcinoma.

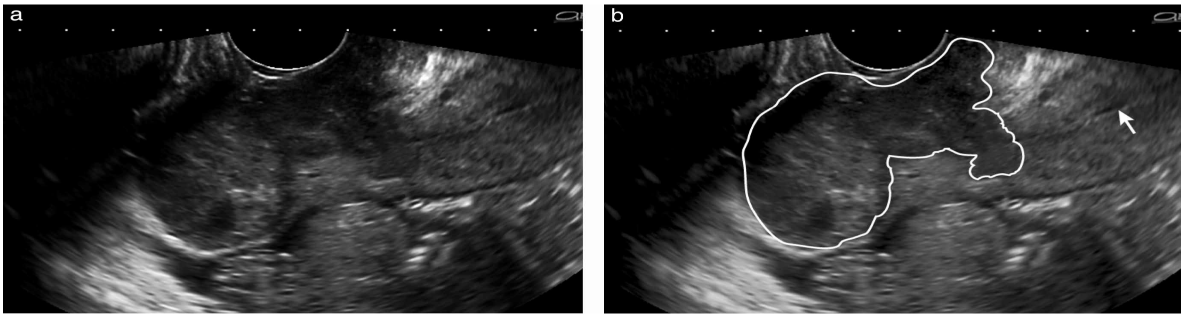

Fig. 4. Two-dimensional grayscale ultrasound image of invasive cervical carcinoma [5]

Magnetic Resonance Imaging (MRI): MRI provides detailed images of the cervix, uterus & surrounding structures. It is beneficial in identifying the extent of cervical cancer and detecting any spread to nearby tissues or organs. MRI is a favored imaging technique for staging cervical cancer. Using powerful magnetic fields and radio waves, this non-invasive technique may take detailed images of the internal organs. The primary cervical cancer tumor's size and location and any adjacent lymph nodes that might be impacted can all be determined using MRI. If cancer has spread to surrounding tissues or organs, it can also assist in figuring that out. A complete image of the extent of cancer can be obtained by combining MRI with additional imaging methods like CT or PET scans. Fig 5 explains how MRI and clinical staging are used in cervical cancer.

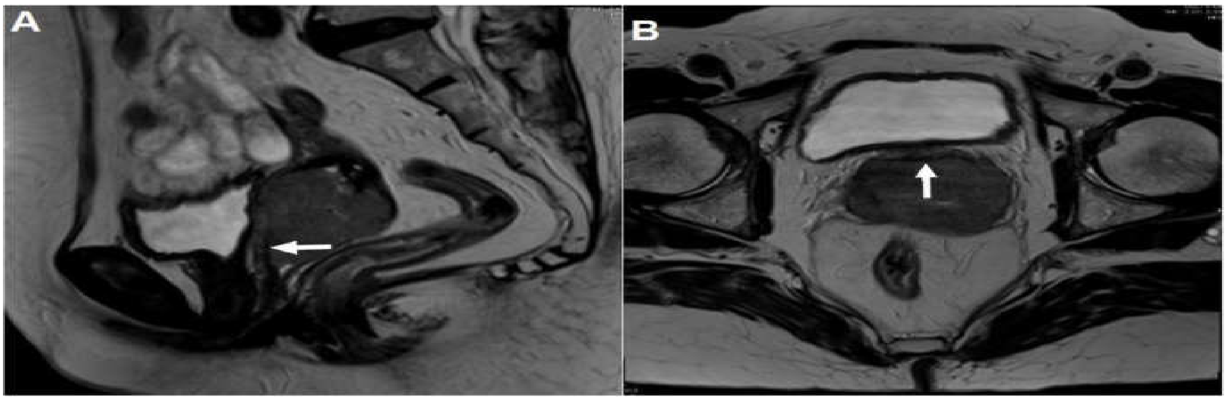

Fig 5. Clinical staging and MRI staging for cervical cancer [6]

Computed Tomography (CT) Scan: A CT scan can also provide images of the cervix, uterus, and surrounding structures. In some circumstances, a CT scan image may aid in diagnosing cervical cancer. However, CT is not frequently the initial imaging modality for detecting cervical cancer. Instead, it is commonly utilized for staging (calculating cancer's size and spread) following a diagnosis. The physical and pelvic examinations will often followed by a Pap smear or HPV test to check for abnormal cells in the cervix if the doctor feels the patient may have cervical cancer. If there are any anomalies, the doctor might suggest more testing, such as a colposcopy (a treatment that examines the cervix under a special microscope) or a biopsy (a surgery to remove a tiny tissue sample). Fig 6 Explains the CT image of Cervical Cancer.

Fig. 6. CT image of cervical cancer [7]

Positron Emission Tomography (PET) Scan: Detect the spread of cervical cancer to other parts of the body. It uses a radioactive tracer to identify cancer cells that are overgrowing. Figure 7 clearly shows what stage II PET seems to be like.

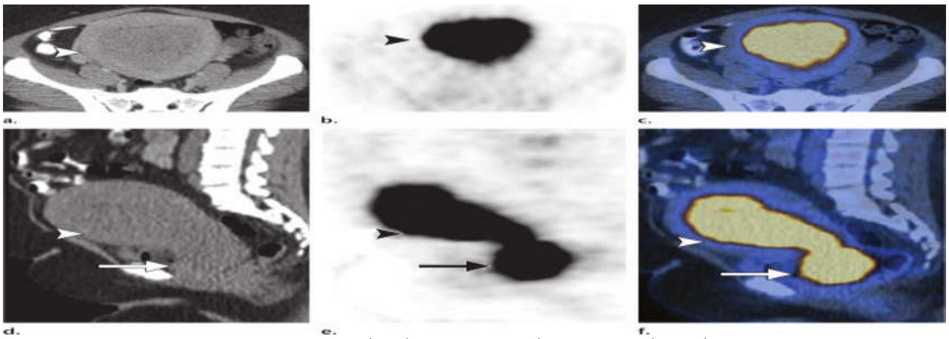

Fig. 7. Stage IIA cervical cancer using PET/CT image [8].

Histopathological Image: In general, a combination of imaging modalities may be used to evaluate cervical cancer comprehensively. Deep learning networks have helped create innovative picture segmentation models with improved performance in recent years. The deep neural networks demonstrated reasonable accuracy rates on various standard datasets—two types of image segmentation techniques: semantic and instance segmentation. Semantic segmentation is a problem of pixel classification. Each pixel in the image with a class in this segmentation process model is detected and delineated via instance segmentation. Deep learning is a critical method of artificial intelligence. Fig 8. Explains what images seem like in the histopathological findings.

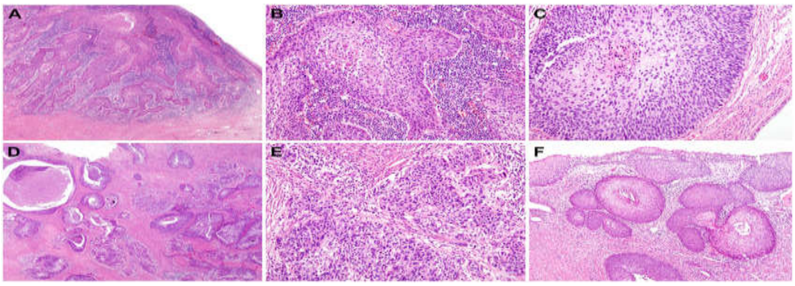

Fig. 8. Histopathological images of cervical cancer [9].

Multiple imaging modalities—such as ultrasound, MRI, CT, PET scans, and histopathological imaging- play crucial roles in detecting, staging, and managing cervical cancer. Each technique offers unique advantages depending on clinical needs, with deep learning further enhancing diagnostic precision. Integrating these methods provides a comprehensive approach for early and accurate cervical cancer evaluation.

2. RELATED WORK

Artificial intelligence is one of the new computer science fields that helps to deal with human intelligence without humans. It is generally accepted that all forms of intelligence result from experience and learning. As a result, it is primitive for the computer to gain intelligence by studying the information accessible within the relevant field. Following the digital revolution's onset, data gathering and storing in medical science received a significant impetus. As a result, many drawbacks should be overcome in the field of ML in medical science or medicine. To prevent cervical cancer, the authors implemented an automatic cervical intraepithelial neoplasia (CIN) classification algorithm for the screening tool. To enhance the robustness of the prediction, they used the gene-assistance module as an optional option. Multiple measurements were accomplished to assess performance. By screening for cervical cancer risk factors and applying an error correction mechanism, erroneous data were identified and corrected during data collection. [10].

2.1 Machine Learning Techniques for Cervical Cancer Diagnosis

ML has emerged as a powerful tool in enhancing early detection, accurate classification, and prognosis prediction in cervical cancer. Traditional diagnostic methods, although valuable, are often subjective, time-consuming, and prone to inter-observer variability. ML techniques address these limitations by automatically analyzing large datasets, uncovering hidden patterns, and assisting healthcare providers in data-driven decision-making. The literature survey briefly overviews the ongoing studies on ML methods used for cervical cancer detection. The review describes about list of different algorithms and strategies employed.

Masatoyo Nakajo 2021 et al.[11] proposed CADx testing, when both offline-trained architectures were run simultaneously on the same platform back-to-back, the system's effectiveness was demonstrated by the results. Six different ML algorithms have been used in cervical cancer patients. Machine learning may help predict tumor progression based on clinical and pretreatment 18F-FDG PET-based radiomic features. The significant advantage of this method is that it is more accurate than the existing one. Still, there is a scope for improvement in the performance evaluation.

H. Zhang 2021 et al.[12] A Raman spectroscopy was used for tissue data collected from five different approaches were used to remove the data background. Other types of feature extraction algorithms have been used. These are some of the feature extraction models used to create many models. S. Zhang 2021 et al.[13] proposed a Synthetic Minority Oversampling Technique (SMOTE) for analyzing Squamous cell carcinoma during a cervical biopsy. In the case of cervical pathology examination, it may lead to time consumption and increase the risk of misdiagnosis or missing diagnosis. Sohely Jahan 2021 et al. [14] presented a machine learning classification technique that can detect cervical cancer and aid in early diagnosis classification algorithms to diagnose accurately. K. Fernandes 2018 et al. [15]. Hence, some data privacy problems are more complicated to carry out. This work mainly concentrated on cervical cancer, which is rapidly increasing in developing countries.

P. Guo et al. 2021 et al. [16] proposed ML models that could determine the likelihood of survival and be specific to the location recurrence in CC and direct personal surveillance. The models can simultaneously predict many outcomes using machine learning techniques, including individual probabilities. K. Kaushik 2022 et al. [17] proposed a method for data analysis by importing the dataset and libraries. Secondly, they performed data standardization and data visualization. After that, the ML model is enabled to detect and diagnose cervical cancer in women. For evaluation, they trained the model and indicated the performance level. The XGBoost technique supported a deep learning model, providing the best performance between numerous frameworks already used in the same field. For training the model, they used a gradient boost algorithm, which is more robust. The significant advantage of this work is accuracy. They needed to work on performance levels. L. Li 2022 et al. [18] proposed after predicting patients with LNM, they made use of MRI features to detect cervical cancer. The employed histology was considered the gold standard, and radionics-MRI was the diagnostic tool for patients with cervical cancer. The overall diagnostic odds ratio (DOR) was examined using MRI-based radiometrics. Ongoing research with strict standardization of radionics processes is required to assess this method's accuracy and diagnostic efficacy before it is further developed as a routine practice in medicine. Naif Al Mudawi 2022 et al. [19] proposed an ML Algorithm for assessing and categorizing cervical cancer to aid doctors in correctly diagnosing the disease. To draw attention to the human papillomavirus, 132 Saudi Arabian volunteers who participated in this study were also surveyed concerning computer-aided cervical cancer prediction (HPV). Deepa. P 2022 et al.[20] proposed a methodology using deep learning and machine learning to predict survival from several databases, and given a thorough analysis of the study, most authors used freely accessible SEER and TCGA data sources. The researchers must prioritize the imputation methods as well. The clinical decision-making process is now significantly aided by machine learning algorithms. Finally, the directions for future research in the survival field were highlighted, along with the research gaps that still need to be filled in. In survival analysis, the statistical subfield focuses on forecasting the timing of events.

D. Luo., 2022 et al.[21] presented a dual-supervised sampling network structure and segmented cell nuclei using image compression of the original images. This method guaranteed accurate segmentation while significantly reducing the convolution calculation required for picture feature extraction. The results of the experiments demonstrated that, in comparison to traditional networks, this technology took both speed and accuracy into account, and it had a promising future in medical imaging.

Table 1. Comparative Analysis of Cervical Cancer Detection Using Machine Learning

Author and Year	Merits	Demerits	Dataset	Accuracy
H. Sartor et al., 2020 [1]	Auto-segmentations by CNN in cervical cancer	No precise method for manual correction	UCI dataset	Accuracy not specified
J. Lima et al., 2021 [2]	Immunotherapy and systemic therapy for cervical cancer	Limited generalizability	---	Accuracy not specified
ICO, 2016 [3]	Comprehensive report on HPV and related diseases	Lack of specific diagnostic methods for cervical cancer	---	Accuracy not specified
PostDICOM, --- [4]	Overview of medical imaging modalities	Lack of direct reference to cervical cancer imaging	---	Accuracy not specified
--- [5]	Study on early-stage cervical cancer detection	The limited scope of the comparison	---	Accuracy not specified
S. Nawapun et al., 2021 [6]	MRI staging for cervical cancer	Clinical staging limitations	MRI data	Accuracy not specified
M. C. Leseduarte and R. Quintanilla, 2015 [7]	Mathematical modeling of material properties	No direct link to cervical cancer detection	---	Accuracy not specified
H. Son et al., 2010 [8]	PET/CT evaluation for cervical cancer	Limited to advanced cancer cases	PET/CT data	Accuracy not specified
G. E. Bae et al., 2016 [9]	High-grade squamous lesions identified	Limited to specific cases of vulvar lymphangioma	---	Accuracy not specified
Y. Yu	Artificial intelligence	No specific dataset	Kaggle	biopsy

2021et al., [10]	prediction of cervical cancer survival	was mentioned	dataset	(98.33%) and cytology (98.65%)
M. Nakajo et al., 2022 [11]	Machine learning-based PET/CT feature evaluation	Focused on pretreatment stages	PET/CT radiomic features	Accuracy not specified
H. Zhang et al., 2021 [12]	Rapid identification of cervical cancer types using Raman spectroscopy	High cost and accessibility concerns	Raman spectroscop y data	96.3%
S. Zhang et al., 2021 [13]	Classification model for cervical tissue staging	Limited applicability to real clinical settings	Pathological tissue images	86.67%.
S. Jahan et al., 2021 [14]	Automated early-stage cervical cancer detection	Requires improvement in model reliability	out-of-bag (OOB) dataset	97%
K. Fernandes et al., 2018 [15]	Automated cervical cancer screening using digital colposcopy	Needs better decision-making transparency	Digital colposcopie s	94%
P. Guo et al., 2016 [16]	Nuclei-based features for cervical cancer histology	A model may be affected by histology variations	Histology image datasets	Accuracy 90%
K. Kaushik et al., 2022 [17]	framework for predicting cervical cancer risk	Needs to improve the overall performance	---	96.5%.
L. Li et al., 2022 [18]	Meta-analysis of MRI radiomic features for metastasis prediction	Focuses on lymph node metastasis prediction	MRI data from various studies	95.3%
N. Al Mudawi and A. Alazeb, 2022 [19]	ML-based cervical cancer prediction model	Algorithm instability and lack of reliability	---	99%
D. P and G. C, 2022 [20]	Review of ML and DL techniques for cancer survival prediction	Focus on generalized survival prediction	SEER dataset and the TCGA	98.4%
D. Luo et al., 2022 [21]	Real-time cervical cell nucleus segmentation	Needs to improve the performance	Tongji Hospital's private	96%

			datasets	
Y. Huang et al., 2022 [22]	Risk factors for cervical lymph node metastasis	Needs improvement in predictive model accuracy	SEER database	Accuracy is about 95%
R. Surendiran et al., 2022 [23]	Application of AI in cervical cancer prediction	Limited real-world application	Herlev dataset	99.5%
T. B. Nguyen et al., 2022 [24]	Identification of crucial genes and pathways for cervical cancer	Focus on inflammatory bowel disease rather than cervical cancer	GEO, TCGA datasets	0.923 AUC
D. Reijtenbagh et al., 2022 [25]	Machine learning in brachytherapy for cervical cancer	Needs further refinement in clinical settings	UMCU dataset	Accuracy not specified
E. Nsugbe, 2022 [26]	Cybernetics for Enhanced Cervical Cancer Care Strategy	Lacks evidence for practical application	(UCI) database	90%+
O. Y. Dweekat and S. S. Lam, 2022 [27]	Integrated PCA and GA for cervical cancer diagnosis	Needs improvement in diagnosis reliability	(UCI) database	Accuracy is about 98.26%
T. Torheim et al., 2017 [28]	Auto-delineation of cervical Cancer using MRI and ML	Limitations in auto-delineation accuracy	MRI data	87%
Z. Yue et al., 2021 [29]	Automatic lesion segmentation using DL	This may lead to errors in segmentation	---	97.3%
M. K. Soumya et al., 2016 [30]	Texture analysis for cervical cancer detection	Limited scope in classification techniques	(UCI) database	87.49%
A. Dongyao Jia et al., 2020 [31]	CNN-SVM network for cervical cancer detection	Needs better feature extraction methods	---	95%

Y. Huang., 2022 et al. [22] used in-build models; machine learning was a potent tool. The SEER database provided information on PTMC, including ten demographic and clinicopathological traits. The significant advantage of this work is to improve the robustness and simplicity and not have any collinearity. R. Surendiran 2022 et al. [23] proposed a novel method that classified and distinguished cervical cytopathology images using DL and ML. According to the review, processing images from cervical cytopathology was an increasing problem. Future models could be more intricate to increase accuracy. In this paper, they have briefly explained

ML and DL technologies used in AI. T.B. Nguyen., 2022 et al. [24], the survival analysis employed to examine hub genes may be biased in cancer research. Twelve miRNAs, five transcription factors, and three hub genes comprise the mRNA regulatory network of microRNAs and transcription factors. A gene-drug interaction study discovered seven pharmaceutical compounds interacting with the hub gene.

D. Reijtenbagh., 2022 et al .[25] proposed a study that determined if patients with cervical cancer could benefit from OVH-based QA in a clinical environment involving multiple institutions for clinical practice. This study investigates the construction of an ML-driven prediction machine using two opposing approaches. This study aimed to develop a novel cybernetic system that demonstrated how humans and machines could work together to create a superintelligence framework that would ultimately enable better clinical care practices. The suggested method enables an automated prioritization of patient care. It provides a more detailed understanding of the extent of cancer without needing a physical examination. The designated block within the cybernetic loop is also anticipated to provide the best care strategy. This study aimed to fill this gap by presenting two opposing prediction machine approaches for monitoring the extent of cervical cancer. The readily accessible factors are transformed via PCA, and the modified features are utilized as input for MLP training. A subset of the original elements is selected instead of the PCA method. Comparing the effectiveness of the PCA-GA-MLP integrated framework to nine alternative algorithmic approaches to classification. As a result, it helps reduce the cervical cancer death rate and improve women's care, which related costs by focusing on preventative measures is considered a significant advantage. Limitations are needed to enhance the performance level [27].

They mainly described how classification is done with the help of MRI to analyze different stages in cervical cancer patients—preprocessed via morphological processes like erosion and dilation combined with contrast enhancement and segmentation algorithms [30]. SVM and LDA classifiers created novel features for segmented epithelial areas' automatic CIN grade classification. It increases the precision of computerized cervical neoplasia diagnosis thanks to the development of enhanced image processing techniques, potentially enabling earlier identification of cancer among women globally. (CNN)-(SVM) models were suggested to categorize the cervical cells effectively, and a new framework was developed based on solid features. CNN combined with the robust features retrieved by the (GLCM) and Gabor. The framework's three critical components were feature extraction, fusion, and classification. On two separate datasets, the approach was assessed [31]

The key finding from the above literature survey is to provide an identification and categorization of cervical cancer. Combining the advantages of several models, techniques such as CNNs for tumor identification and picture segmentation and ensemble learning approaches have demonstrated efficacy in enhancing classification outcomes. Moreover, the application of radiomics and feature extraction offers essential insights into the fundamental traits of cervical cancer, enabling more accurate prognosis and risk assessment. Predictive modeling has made an additional contribution by providing instruments to predict survival, recurrence, and metastasis, all of which are essential for individualized treatment plans. All things considered, integrating this state-of-the-art technology not only speeds up the diagnostic procedure.

3. METHODOLOGY OF MACHINE LEARNING CLASSIFICATION

The proposed workflow consists of the following steps: Data collection, Data preprocessing, Feature Extraction, Model selection, Model training, Model evaluation, Model optimization, and Deployment. In the data collection step, cervical cancer datasets containing medical images and clinical data are collected from various sources, such as hospitals, research institutions, and public repositories. The collected data are preprocessed to remove noise, artifacts, and other unwanted features in the data preprocessing step. This step may involve image resizing, normalization, and augmentation techniques to increase the data's quality and volume. In the feature extraction step, texture, shape& color features are extracted from the preprocessed data using various algorithms and techniques such as histogram-based methods, wavelet transforms, and local binary patterns.

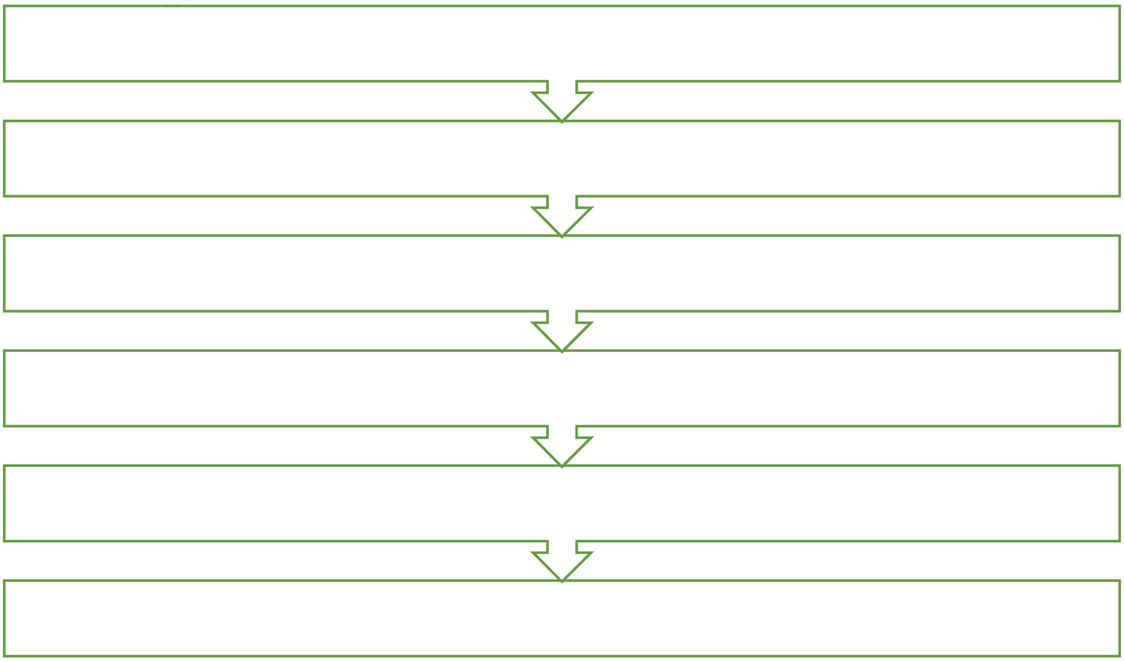

Fig 9: (Computer Aided System) CAD System

In the model selection step, we select ML models such as CNN, SVM, and decision trees. The choice of model depends on the specific task and dataset. We train selected models using the preprocessed and extracted features in the model training step. This step involves setting various hyperparameters, defining the loss function, and optimizing the model using backpropagation and gradient descent techniques. In the model evaluation step, we evaluate the trained model using a separate dataset or through cross-validation.

In the model optimization step, we further optimize the model using hyperparameter tuning, regularization, and ensemble methods to improve its performance. In the deployment step, we deploy the optimized model in a clinical setting for automated cervical cancer detection, diagnosis, or prediction. This step involves integrating the model with existing clinical workflows and systems and ensuring its safety, efficacy, and usability. Fig 9. Briefly explain the Block diagram of the (Computer Aided System) CAD System.

4. CERVICAL CANCER IMAGE DATASET

The Cervix Image Database features colposcopy images and key metadata like patient age, number of pregnancies, and HPV status, all aimed at aiding cervical cancer detection. Similarly, the Kaggle Cervical Cancer Screening Dataset provides labeled cervical cell images, distinguishing between normal and abnormal cells to support automated screening efforts. Meanwhile, the SIPaKMeD dataset offers cytology images sorted into categories of normal and pre-cancerous or cancerous cells, assisting in diagnostic processes. The Colposcopy Image Dataset includes images and videos of the cervix, each annotated with details of cervical abnormalities (CIN) to aid in cancer detection. Lastly, the Cervical Histology Dataset consists of histology slide images classified according to cancer type and grade, helping advance automated diagnosis and prediction models. The table below describes the type and number of images in that dataset, followed by the splitting of cancer cells and their applications.

Table 2 Dataset of Cervical Cancer

Dataset Name	Images	Image Types	Splitting (Cancer / Pre-cancer / Non-cancer)	Applications
Cervix Image Database	4000+ images	Colposcopy images + Metadata (age, parity, HPV status)	Labels: Normal, CIN1 (mild dysplasia / pre-cancer), CIN2, CIN3 (severe dysplasia / cancer)	Automated cervical cancer detection
Kaggle Cervical Cancer Screening Dataset	8000+ images	Cytology (Pap smear) cell images	Normal, LSIL (Low-grade Squamous Intraepithelial Lesion - pre-cancer), HSIL (High-grade - cancer)	Automated cervical cancer screening
SIPaKMeD Cervix Cytopathology Dataset	4049 individual cell images	Cytology images (Pap smear)	5 classes: Superficial-Intermediate (normal), Parabasal (normal), Koilocytotic (pre-cancerous), Dyskeratotic (pre-cancerous/cancerous), Metaplastic (normal)	Cell-based cervical cancer diagnosis
Colposcopy Image Dataset	800 images and videos	Colposcopy images/videos	Labels for Normal, Pre-cancer (CIN1, CIN2), and Cancer (CIN3+)	ML-based detection and classification of cervical abnormalities

Cervical Histology Dataset	1000+ histology slide images	Histopathology images	Labels: Normal, CIN1, CIN2, CIN3, Invasive Cancer	Cancer grade diagnosis and prediction

5. PERFORMANCE METRICS

For predictions of the machine learning model, it is categorized into further classification and regression. Classification is a prediction type that organizes similar attributes and gives the output variables. Some classification metrics are accuracy, precision, recall, and F1 score. The opposite of categorization is regression. A prediction method known as regression uses a numerical outcome instead of a categorization variable to make the forecast.

Accuracy: Accuracy measures the overall correctness of a system's output. In medical image processing, accuracy measures how well a system's production matches the ground.

(1)

Precision: Measures the degree of agreement between a system's output and the ground truth. In medical image processing, precision measures whether a system's result is consistent across multiple runs.

(2)

Recall: In medical image processing, recall measures the degree to which a system's result is consistent across multiple images.

(3)

F1 score: In medical image processing, the F1 score is often used to evaluate the overall accuracy of a system's output.

(4)

MSE (Mean Squared Error): MSE (Mean Squared Error) can be used in medical image processing to evaluate how well an image reconstruction algorithm performs.

(5)

6. CONCLUSION

One of the most promising and quickly increasing fields of machine learning is diagnosing medical data. Cervical cancer is one of the deadliest cancers, but if caught early enough, it is curable. As a result, developing a computerized method for cervical cancer diagnosis is critical. Our survey findings indicate that machine learning may help predict and detect cervical cancer. Age, smoking status, and HPV status are just a few of the significant factors our analysis revealed to be highly related to the risk of cervical cancer. These results underline how crucial early detection and prevention efforts are, especially for high-risk populations. Our research significantly impacts the creation of ML algorithms for cervical cancer screening. Our study will aid in the development of more precise and effective screening tools by identifying the most predictive features.

In conclusion, automated cervical cancer identification and diagnosis in medical imaging. Additional study is required to create trustworthy algorithms that can be used in clinical practice.

Future research should enhance these algorithms' clinical utility and generalizability and test how well they function in practical situations. To advance the development of more precise, effective, and widely available cervical cancer detection and diagnosis techniques, we believe this survey will be an invaluable resource for researchers and clinicians working in this area.

REFERENCES

[1] H. Sartor *et al.*, "Auto-segmentations by convolutional neural network in cervical and anorectal cancer with clinical structure sets as the ground truth," *Clin. Transl. Radiat. Oncol.*, vol. 25, pp. 37–45, 2020, doi: 10.1016/j.ctro.2020.09.004.

[2] J. Lima, Z. Ali, and S. Banerjee, "Immunotherapy and Systemic Therapy in Metastatic/Recurrent Endometrial and Cervical Cancers," *Clin. Oncol.*, vol. 33, no. 9, pp. 608–615, 2021, doi: 10.1016/j.clon.2021.07.002.

[3] ICO, "Human Papillomavirus and Related Diseases Report," no. October, 2016.

[4] "Medical Imaging Types and Modalities | PostDICOM."

[5] "2011 -Early stage cervical cancer agreement between ultrasound and.pdf."

[6] S. Nawapun, C. Aphinives, W. Srisitthipropha, K. Thamronganantasakul, and A. Temtanakitpaisan, "Correlation of clinical staging and MRI staging for cervical cancer," *Egypt. J. Radiol. Nucl. Med.*, vol. 52, no. 1, 2021, doi: 10.1186/s43055-021-00544-8.

[7] M. C. Leseduarte and R. Quintanilla, "Lower bounds of end effects for a nonhomogeneous isotropic linear elastic solid in anti-plane shear," *Math. Mech. Solids*, vol. 20, no. 2, pp. 140–156, 2015, doi: 10.1177/1081286514544256.

[8] H. Son *et al.*, "PET/CT Evaluation of cervical cancer: Spectrum of disease," *Radiographics*, vol. 30, no. 5, pp. 1251–1268, 2010, doi: 10.1148/rg.305105703.

[9] G. E. Bae, G. Yoon, Y. J. Song, and H. S. Kim, "High-grade squamous intraepithelial lesion arising adjacent to vulvar lymphangioma circumscriptum: A tertiary institutional experience," *Oncotarget*, vol. 7, no. 30, pp. 48120–48129, 2016, doi: 10.18632/oncotarget.10158.

[10] "2021Novel artificial intelligence machine learning approaches to precisely predict survival and site-specific recurrence in cervical cancer A multi-institutional study.pdf."

[11] M. Nakajo *et al.*, "Machine learning based evaluation of clinical and pretreatment 18F-FDG-PET/CT radiomic features to predict prognosis of cervical cancer patients," *Abdom. Radiol.*, vol. 47, no. 2, pp. 838–847, 2022, doi: 10.1007/s00261-021-03350-y.

[12] H. Zhang *et al.*, "Rapid identification of cervical adenocarcinoma and cervical squamous cell carcinoma tissue based on Raman spectroscopy combined with multiple machine learning algorithms," *Photodiagnosis Photodyn. Ther.*, vol. 33, p. 102104, 2021, doi: 10.1016/j.pdpdt.2020.102104.

[13] S. Zhang *et al.*, "Research on application of classification model based on stack generalization in staging of cervical tissue pathological images," *IEEE Access*, vol. 9, pp. 48980–48991, 2021, doi: 10.1109/ACCESS.2021.3064040.

[14] S. Jahan *et al.*, "Automated invasive cervical cancer disease detection at early stage through suitable machine learning model," *SN Appl. Sci.*, vol. 3, no. 10, 2021, doi: 10.1007/s42452-021-04786-z.

[15] K. Fernandes, J. S. Cardoso, and J. Fernandes, "Automated Methods for the Decision Support of Cervical Cancer Screening Using Digital Colposcopies," *IEEE Access*, vol. 6, pp. 33910–33927, 2018, doi: 10.1109/ACCESS.2018.2839338.

[16] P. Guo *et al.*, "Nuclei-Based Features for Uterine Cervical Cancer Histology Image Analysis with Fusion-Based Classification," *IEEE J. Biomed. Heal. Informatics*, vol. 20, no. 6, pp. 1595–1607, 2016, doi: 10.1109/JBHI.2015.2483318.

[17] K. Kaushik *et al.*, "A Machine Learning-Based Framework for the Prediction of Cervical Cancer Risk in Women," *Sustain.*, vol. 14, no. 19, 2022, doi: 10.3390/su141911947.

[18] L. Li *et al.*, "A meta-analysis of MRI-based radiomic features for predicting lymph node metastasis in patients with cervical cancer," *Eur. J. Radiol.*, vol. 151, no. December 2021, p. 110243, 2022, doi: 10.1016/j.ejrad.2022.110243.

[19] N. Al Mudawi and A. Alazeb, "A Model for Predicting Cervical Cancer Using Machine Learning Algorithms," *Sensors*, vol. 22, no. 11, 2022, doi: 10.3390/s22114132.

[20] D. P and G. C, "A systematic review on machine learning and deep learning techniques in cancer survival prediction," *Prog. Biophys. Mol. Biol.*, vol. 174, no. June, pp. 62–71, 2022, doi: 10.1016/j.pbiomolbio.2022.07.004.

[21] D. Luo *et al.*, "Dual supervised sampling networks for real-time segmentation of cervical cell nucleus," *Comput. Struct. Biotechnol. J.*, vol. 20, pp. 4360–4368, 2022, doi: 10.1016/j.csbj.2022.08.023.

[22] Y. Huang, Y. Mao, L. Xu, J. Wen, and G. Chen, "Exploring risk factors for cervical lymph node metastasis in papillary thyroid microcarcinoma: construction of a novel population-based predictive model," *BMC Endocr. Disord.*, vol. 22, no. 1, pp. 1–13, 2022, doi: 10.1186/s12902-022-01186-1.

[23] R. Surendiran, M. Thangamani, S. Monisha, and P. Rajesh, "Exploring the Cervical Cancer Prediction by Machine Learning and Deep Learning with Artificial Intelligence Approaches," *Int. J. Eng. Trends Technol.*, vol. 70, no. 7, pp. 94–107, 2022, doi: 10.14445/22315381/IJETT-V70I7P211.

[24] T. B. Nguyen, D. N. Do, M. Le Nguyen-Thi, H. Hoang-The, T. T. Tran, and T. Nguyen-Thanh, "Identification of potential crucial genes and key pathways shared in Inflammatory Bowel Disease and cervical cancer by machine learning and integrated bioinformatics," *Comput. Biol. Med.*, vol. 149, no. April, p. 105996, 2022, doi: 10.1016/j.compbiomed.2022.105996.

[25] D. Reijtenbagh *et al.*, "Multi-center analysis of machine-learning predicted dose parameters in brachytherapy for cervical cancer," *Radiother. Oncol.*, vol. 170, pp. 169–175, 2022, doi: 10.1016/j.radonc.2022.02.022.

[26] E. Nsugbe, "Towards the use of cybernetics for an enhanced cervical cancer care strategy," *Intell. Med.*, vol. 2, no. 3, pp. 117–126, 2022, doi: 10.1016/j.imed.2022.02.001.

[27] O. Y. Dweekat and S. S. Lam, "Cervical Cancer Diagnosis Using an Integrated System of Principal Component Analysis, Genetic Algorithm, and Multilayer Perceptron," *Healthc.*, vol. 10, no. 10, 2022, doi: 10.3390/healthcare10102002.

[28] T. Torheim *et al.*, "Autodelineation of cervical cancers using multiparametric magnetic resonance imaging and machine learning," *Acta Oncol. (Madr).*, vol. 56, no. 6, pp. 806–812, 2017, doi: 10.1080/0284186X.2017.1285499.

[29] Z. Yue, S. Ding, X. Li, S. Yang, and Y. Zhang, "Automatic Acetowhite Lesion Segmentation via Specular Reflection Removal and Deep Attention Network," *IEEE J. Biomed. Heal. Informatics*, vol. 25, no. 9, pp. 3529–3540, 2021, doi: 10.1109/JBHI.2021.3064366.

[30] M. K. Soumya, K. Sneha, and C. Arunvinodh, "Cervical cancer detection and classification using texture analysis," *Biomed. Pharmacol. J.*, vol. 9, no. 2, pp. 663–671, 2016, doi: 10.13005/bpj/988.

[31] A. Dongyao Jia, B. Zhengyi Li, and C. Chuanwang Zhang, "Detection of cervical cancer cells based on strong feature CNN-SVM network," *Neurocomputing*, vol. 411, pp. 112–127, 2020, doi: 10.1016/j.neucom.2020.06.006.