Journal of Theoretical and Applied Information Technology
15st June 2026. Vol.104. No.11

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AI-ENABLED CERVICAL CANCER RISK PREDICTION USING YOLOV12 DEEP FEATURES AND OPTIMIZED RF-XGBOOST CLASSIFICATION

Dr. S. FOUZIA SAYEEDUNNISA[1] , Dr. MANIZA HIJAB[2], Dr.S.GOMATHI[3] , MYLAVARAPU KALYAN RAM[4], SHAIK MABASHA[5], Dr.GARLAPATI NARAYANA[6] , Dr.VANKUDOTHU MALSORU[7], Dr. SIVA KUMAR PATHURI[8]

[1]Associate Professor, Information Technology Department , Muffakham Jah College of Engineering and Technology, Banjara Hills, Hyderabad, Telangana, India
[2]Associate Professor,Department of Computer Science and Engineering, Muffakham Jah College of Engineering and Technology, Banjara Hills, Hyderabad. Telangana, India.
[3]Assistant Professor, Electronics Engineering (VLSIDT), K.S.Rangasamy College of Technology, Tamil Nadu, India.
[4]Assistant Professor Department of CSE, Aditya University, Surampalem, Andhra Pradesh-533437, India.
[5]Assistant Professor, Department of CSE (AIML&IoT) , VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, 500090
[6]Associate professor, Department of AIML, Chaitanya Bharathi Institute of Technology, Gandipet, Hyderabad, India.
[7]Associate Professor, CSE(AI&ML) Dept. CMR Technical Campus, Kandlakoya, Hyderabad, Medchal, Telangana -501401, India.
[8]Professor, Dept of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, AP, India.

Email: [1]fouzia@mjcollege.ac.in, [2]manizahijab@mjcollege.ac.in, [3]gomathi@ksrct.ac.in, [4]kalyanram1985@gmail.com, [5]mabashapec@gmail.com, [6]narayanag.1973@gmail.com, [7]malsoru.it@cmrtc.ac.in, [8]spathuri@kluniversity.in

ABSTRACT

Cervical cancer is the second malignancy that is highly diagnosed among women and is regarded as one of the preventable ones. However, despite the existence of effective prevention measures, the literature has shown that the level of ignorance is high concerning cervical cancer, its causative factors, and prevention measures in the developing countries, in particular. The medical students and health care professionals are in a strategic position to raise the awareness of people by elevating the level of their knowledge regarding the symptoms, risk factors, and the early screening practices. Approximately 500,000 new cases of cervical cancer are reported each year all over the world with over 300,000 deaths. The main etiological factor leading to the development of the disease is the persistent infection with the human papillomavirus (HPV) of high-risk subtypes. Although cervical cancer cases can be considerably avoided due to the organized screening and HPV immunization campaigns, almost 90 percent of the cases occur in the states with low and middle income due to the lack of access to systematic screening and immunization. By contrast, the cervical cancer incidence and mortality rates in high-income countries have decreased tremendously during the last thirty years due to the complex screening and early intervention efforts. Also, fertility saving surgical procedures are now the norm in the treatment of women with cervical cancer at the early and low risk stages. The current paper introduces an advanced ensemble-based deep learning framework that could be utilized to identify and forecast at an early stage cervical cancer risk using both YOLOv12 and RF+XGboost to extract and classify deep features respectively. The suggested YOLOv12-RFXGBoost model has performed better than the conventional classifiers, with the peak accuracy of 96.5, and can be used in the early detection of risks and the provision of clinical decisions. And when compared with some base classifiers like Decision Tree and Support Vector Machine the proposed classifier has given the best accuracy in predicting/detecting cervical cancer.

Keywords: *Cervical cancer, Risk factors, screening, (HPV), YOLOv12, Random Forest, XGBoost, Decision Tree, Support Vector Machine.*

1. INTRODUCTION

The second malignancy that has turned out to be the most dominant cancer affecting women all over the world is cervical cancer, which remains to be a significant threat to the health of women. Its clear and well-developed ethology has made it possible to develop exhaustive prevention and control measures. On its part, the World Health Organization (WHO) made a global appeal to eradicate cervical cancer in May 2018, which was endorsed by over 70 countries and international academic institutions. Nevertheless, WHO[1] gave official publicity to a global strategy on 17 November 2025 to accelerate the eradication of cervical cancer as a cancer-related morbid and mortality cause among women, with 194 countries committing to take joint action towards this outcome. Reports by the WHO and International Agency in Research on Cancer (IARC) show that an estimated 600,000-930,000 new cases were reported in the world in 2020 with close to half of the affected women dying of the disease[2]. Cervical cancer is the fourth most predominant cancer that causes death in women across the globe, and close to 88 percent of them are killed in low- and middle-income nations. The mortality rate in such areas is almost 20 percent more than in the high-income countries and this is mostly because of the lack of accessibility to the organized screening, early diagnosis and vaccination programs. Worryingly, the occurrence of cervical cancer amongst young women has gone on to rise significantly in recent years as compared to the levels of about 25 to 45 even in the developed countries where there is a relatively high awareness level. Geographical inequality further brings in the severity of the disease[3]. The highest prevalence and mortality rates are observed in Africa, and the rates in North America, Australia/New Zealand, and Western Asia are 710 times less. Cervical cancer is the second type of cancer that is mixed up in China among women and national statistics show that over the last twenty years the rate of incidence and mortality have been on the increase. These statistics highlight the importance of better strategies of early detection and risk assessment. The pathogenesis of cervical cancer is closely related to an ongoing infection by high-risk human papillomavirus (HPV)-16, 18, 31, 33 and 45. Some of these risk factors are low socioeconomic status, early marriage, early entry into the sexual realm, having numerous sexual partners, high parity, smoking, and inadequate access to healthcare services. Cervical cancer has a biological origin in the transformation zone of the cervix where there is a cross over between the glandular and squamous epithelial cells. Normal cervical cells develop over a slow transition process of precancerous lesions to invasive cancer, thereby making it possible and essential to detect them at an early stage[4]. Notably, cervical cancer is very preventable and treatable when the disease is detected early, but ignorance, stigma, and poor screening procedures especially among the low socioeconomic women in society are still a key obstacle to this issue. As a response to such problems, this paper suggests a new advanced deep learning-based method of detecting and predicting the risk of cervical cancer at an early stage. In the proposed method, the automated deep features are extracted using YOLOv12 because it is better than the other models in extracting spatial and context features of medical images. These features are then categorized with the help of the ensemble machine learning algorithms, namely, the Random Forest (RF) and the Extreme Gradient Boosting (XGBoost) in order to improve the effectiveness and the predictive power. With the aim of proving the usefulness of the suggested YOLOv12-based framework, a detailed comparative study is made with the reference to the previously proven deep learning models, such as Decision Tree and Support Vector Machine, combined with the same RF and XGBoost classifiers. The experimental data prove that the YOLOv12 + RF/XGBoost hybrid architecture is always better in accuracy, sensitivity, and overall predictive behaviour compared to the Decision Tree- and Support Vector Machine-based one. The presented model has a higher ability to identify diseases in the early stage and is therefore a potential clinical decision support tool as well as a large-scale screening system. This research combines powerful deep learning and ensemble classification methods to help enhance the situation in the area of early diagnosis, death reduction[5], and other methods that could help to eliminate cervical cancer in the world. Figure 1 shows Cervical cancer development image & Detailed vector illustration with uterus and cervix carcinoma stages.

Journal of Theoretical and Applied Information Technology
15st June 2026. Vol.104. No.11

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

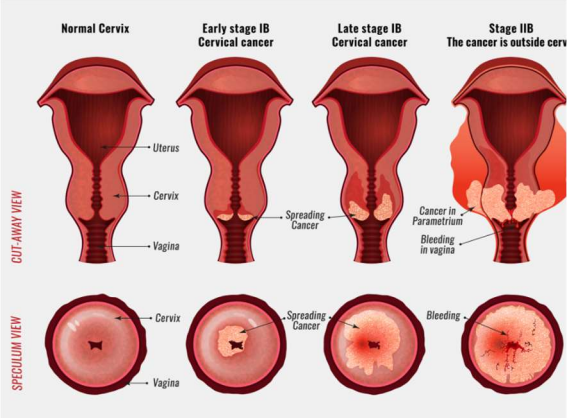

Figure. 1 Cervical Cancer Development Image & Detailed Vector Illustration With Uterus And Cervix Carcinoma Stages.

2. LITERATURE SURVEY

It was reported that awareness of cervical cancer is 70.8-94.0% and awareness of human papillomavirus (HPV) is significantly lower and is not more than 5.6-27.3. The female students were observed to be much more aware than the male students and students in the higher level of education were much more aware in comparison to students with low educational background. A survey done on the 17 and 18 year old teens in Greece indicated that there were no significant differences in awareness about geographical location, i.e. whether living in an urban or a rural setting. The literature evidence suggests that the marketing of cervical cancer is becoming more and more apparent among younger women who are mostly aged 17-30 years, which can be mostly attributed to the lack of knowledge about the cervical cancer, its risk factors and preventive measures. Even though screening, vaccination, and therapeutic interventions have been applied by various researchers to lessen the burden caused by cervical cancer, full eradication of the disease has not yet been realized. Other medical practices such as homeopathy have been considered but the medical practices are questionable and have no enough clinical studies to prove their effectiveness. The conventional medical practice identifies cervical cancer as a complicated illness because of its invasive developmental patterns, possibility of remote metastasis, and complications related to it and early diagnosis is extremely significant. Table 1 provides a comparative introduction into the past studied works where the wide range of methodologies and techniques introduced is used to detect cervical cancer, determine the risk, and manage the condition. This comparative study highlights the necessity of more precise, scaled, and fact-based methods to help in early diagnosis and prevention methods. Table 1 shows the related work for cervical cancer.

Year	Author(s) / Study	Approach / Model	Dataset / Modality	Key Contribution / Results
2025	Okyay et al. Evaluating Cervical Cancer Risk Using ML	Ensemble ML (RF, LightGBM, SMOTE)	UCI Cervical Cancer Risk Factors	High accuracy (~97%) for risk prediction with improved recall after SMOTE balancing.
2023	Guo et al. Deep Learning-Based Cervical Cytology Screening	Systematic review of DL methods	Cytology image datasets	Comprehensive survey on DL approaches for cervical screening and detection.
2025	Li et al. Preoperative & Postoperative Cervical Cancer Prediction	Multicenter ML model integrating MRI + clinicopathological features	MRI + Clinical records	Combined imaging and clinical data for prognosis and prediction with ML. (
2025	Jadhav et al. DL Based Classification of Cancer Stages	Transfer learning (DL models)	Clinical image datasets	Transfer DL for multi-stage classification of cervical cancer.
2025	Jayasundari & Arumugam ML Approach for Cervical Cancer	ML (LR, RF, SVM, XGBoost)	Risk factor dataset	XGBoost showed superior performance and ROC AUC for early prediction.

Year	Author(s) / Study	Approach / Model	Dataset / Modality	Key Contribution / Results
	Prediction			
2025	JAIT 2025 Explainable AI for Cervical Cell Classification	VGG16 + XAI methods (Grad-CAM, LRP)	Herlev dataset	Achieved ~91.94% accuracy; XAI enhanced interpretability.
2025	Scientific Reports Deep Transfer Learning Comparison	16 pre-trained DL models (ResNet, VGG, DenseNet)	Herlev & Sipakmed	ResNet50 achieved ~95% accuracy; VGG16 ~99.95% on pap smear classification.
2025	Frontiers in Oncology AI in Cervical Pathological Diagnosis	Systematic review	Histopathology datasets	Highlights CNNs achieving 92–98% diagnostic accuracy; implementation challenges discussed.
2024	International Journal RAS Cervical Cancer DL Review	Review of DL cytology & imaging	Mixed imaging	Summarizes limitations, dataset issues, and future directions in DL screening.
2024	Arumugam ML Cervical Cancer Risk Analysis	XGBoost & RF	Biopsy + clinical data	Demonstrated superior predictive performance for risk modeling.
2025	Neural Processing Letters Cervical Cancer DL Classification	CNN DL model	Colposcopy images	Showed reliable classification performance with deep features.

3. PROBLEM STATEMENT

Cervical Cancer is one of the most common cancer-related causes of morbidity and mortality among female population worldwide especially in LMIC countries, where there is less access to organized screening and diagnostic facilities. Despite the fact that the disease is very preventable and treatable, if detected early, there is still insufficient awareness, screening infrastructure, timely diagnosis and inadequate training of healthcare workers, all of which are contributing to poor clinical outcomes. Labor-intensive, lengthy, and variable diagnostic methods like Pap smear testing, HPV testing[6], and colposcopy may not achieve the same level of diagnostic consistency and reliability. In recent years, the success of machine learning and deep learning techniques has led to exciting new prospects for automated diagnosis and risk prediction of cervical cancer using medical imaging data. Most current methods, however, use individual classifiers or manually designed feature extraction methods which are ineffective in extracting complex spatial, contextual, and discriminative patterns in the images of cervical cells. Moreover, the traditional classifiers, like Decision Tree and Support Vector Machine[7] are often limited in generalization, computation cost, overfitting and dataset-inconsistency problems in heterogeneous dataset. These difficulties underscore the importance of a powerful, scalable, and highly precise automated platform for better early detection and risk stratification of cervical cancer.The proposed research aims to tackle these challenges by using an advanced deep feature extraction model, YOLOv12, to efficiently extract high-level discriminative feature representations from cervical medical images and combining them with an ensemble learning approach using RF-XGBoost to enhance the accuracy of cervical cancer risk analysis. The deep features extracted are then classified by adopting a hybrid ensemble learning method combining the Random

Forest and XGBoost methods to increase the prediction robustness, improve the classification accuracy, and lessen the diagnostic errors. The primary challenges of the present work are the absence of an efficient and generalized automated cervical cancer diagnostic system that can provide high precision and reliable risk stratification with lesser computational complexity[8]. The envisioned hybrid architecture can enable an intelligent clinical decision support system for early detection of cervical cancer, which can help in a timely treatment planning for better survival outcomes for the patients.

4. METHODOLOGY

With approximately 9.6 million deaths from cancer worldwide in 2019, it is the second most common cause of death. Cancer is brought on by normal cells undergoing a multi-stage process that ultimately results in a malignant tumour. However, if cancer is detected earlier, it is more likely to respond to the right treatment, increasing the likelihood of survival, reducing morbidity, and requiring less expensive therapy[9]. The complex ecosystem produced by screening and diagnosis techniques is now difficult to analyse from the perspective of a computer-aided diagnosis (CAD) system. Due to a lack of computing resources, these complicated problems are getting worse in many developing countries. Finding the best screening strategy and calculating risk are the main diagnosis issues for all patients who skip the routine screening. Most screening techniques are correlated with the doctor's judgement and experience. One can use the survey to identify the riskiest group and eliminate pointless screening. A plan based on the cancer risk aids in resolving cancer-related problems. According to a recent World Health Organization survey, cervical cancer is the "fourth most prevalent type of cancer."[10] This cancer is risky, particularly in comparison to other cancers. One such cancer is brought on by contracting the HPV virus first [11]. Many researchers found that sexual contact is the main way in which the HPV virus is spread. There are many different types of HPVs, and pattern 18 and category 16 have been linked to cancer. Category 6 and category 11 have been regarded as significant HPVs because they cause cystitis on the surface. These are the highest HPVs because they cause cancer cell tissues in the area. In the past, a neural network was discovered to be an effective and efficient detection algorithm. For various linear models, the researchers presented an ensembled approach and discussed its applicability in various situations. This experiment has produced fruitful findings. Other methods for cancer detection have been investigated, including improved genetic algorithms, ANN, and hierarchical clustering. The authors [12] classified the cancer dataset, and the results indicated that performance varied between 80% and 90% approximately. The authors used various data mining methods and classifiers in 2016 to forecast heart diseases. The range of performance parameters between approximately 45% and 91% has been presented by the researchers. A comparative analysis of various machine learning models used for the early diagnosis of heart disease was conducted in 2017 by the authors. So, the ensembled algorithm approach is shown in Figure 2. The proposed approach, named “An Advanced Deep Feature Extraction and Ensemble Learning Approach Using YOLOv12 and RF-XGBoost for Cervical Cancer Risk Analysis,” enhanced the process of early detection of cervical cancer by combining deep learning with ensemble classification. The unintended positive side effects were improvements in feature generalization, robust performance in the presence of class imbalance and noise, and the avoidance of overfitting because of the ensemble of the RF-XGBoost decision mechanism[13]. A challenge, albeit minor, was the complexity of training and the computational expense. The proposed model demonstrated higher accuracy, F1 score, predictive stability, and diagnostic reliability than other existing models such as SVM, Decision Tree, and standalone CNN-based models for deep feature extraction and clinical decision support using efficient YOLOv12.The proposed model outperforms the SVM, Decision Tree and standalone CNN-based model in terms of accuracy, F1 score, predictive stability and diagnostic reliability using efficient YOLOv12[14] for deep feature extraction and clinical decision support.

Journal of Theoretical and Applied Information Technology
15st June 2026. Vol.104. No.11

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

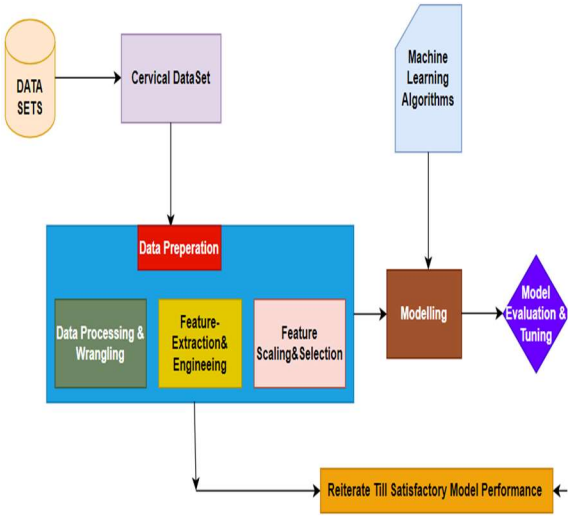

Figure 2: Shows The Architecture Of The Proposed Model.

The proposed research methodology will be structured into four key elements, which include: data acquisition, data preprocessing, predictive model selection (PMS) and model training and evaluation. As illustrated in the figure 1, the methodology is organized in four consecutive steps, which carry out a particular and necessary task in the cervical cancer prediction pipeline. The data acquisition step is aimed at the acquisition of the necessary clinical and demographic information needed in the development of the model. The preprocessing step of data processing deals with the cleaning of the data, missing data, normalization, and feature preparation to fit data in the machine learning algorithms[15]. The predictive model selection (PMS) stage entails the selection of appropriate machine learning and ensemble learning models in order to make appropriate cervical cancer predictions. Lastly, the model training and assessment stage runs the chosen algorithms with the help of a Python-based pipeline and measures their prediction capabilities with the help of standard validation methods This paper discusses various predictive models such as Decision Tree[16], Support Vector Machine and a Cervical Predicting Boosting Algorithm (CPBA) to determine their performance in the diagnosis of cervical cancer. These models are compared and discussed in the following sections.

4.1. Kaggle Dataset Description

The data set in this research was retrieved in the Kaggle repository[17], database under the title Mendeley LBC Cervical Cancer. The data is anonymized medical, demographic, and lifestyle data gathered on women patients. There are numerous missing values in the dataset, which requires thorough preprocessing due to privacy issues and non-answering.

The last set of data include 2000 female patient records each record in the form of image beig a risk factor. The dataset is popular in the prediction of cervical cancer, and it serves as a good standard reference to test machine learning models. And the process flow of the model is shown in Figure 3. A visual representation of the dataset attributes and structure is shown in Figure 4.

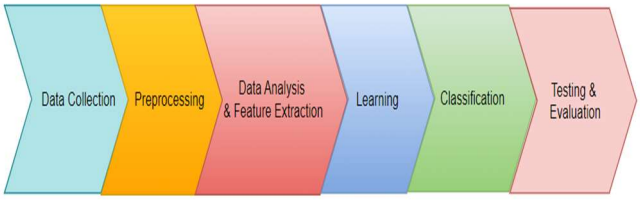

Figure 3: Shows SVM Classifier For Weed Detection

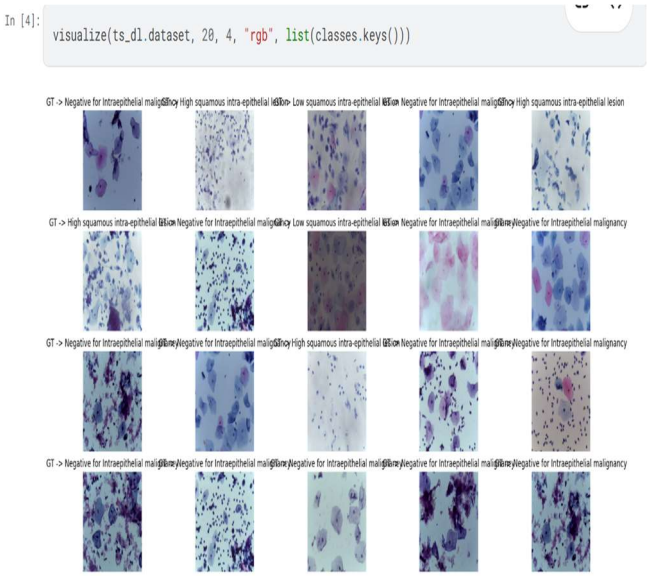

Figure 4: Shows The Risk Factors Of Cervical Cancer Dataset.

4.2. Data Preprocessing:

The data pre-processing is an essential step of the proposed framework because the quality of the input images directly determines the performance, robustness, and reliability of the deep learning-based models. The image pre-processing pipeline in this study is divided into three major phases, including image cleansing, image transformation, and feature extraction and data reduction. Image cleaning is done to overcome the usual image related defects that include noise, artifacts, low contrast and uneven illumination. The dataset of cervical cancer images that was gathered on Kaggle demonstrates significant difference in the quality of the images as they were acquired under different circumstances. The noise suppression and enhancement methods were used to reduce these problems. Median filtering was used to eliminate impulsive noises without destroying key structural content and contrast enhancement methods were used to enhance the appearance of structures with diagnostic significance. Image conversion was performed to make the raw images standardized to be used in deep learning structures. To provide consistency of all images in the dataset, all images were resized into a fixed resolution. MinMax[18] scaling that normalized the pixel intensity was also used to bring pixel values to a homogenous range which improved numerical stability and model convergence speed. Also, data augmentation methods, such as rotation, horizontal and vertical flipping, scaling, and translation, have been added to artificially increase the dataset and decrease the classes imbalance and overfitting. The transformations allow the model to learn strong and invariant cervical tissue patterns representations. The YOLOv12 architecture was used to carry out data reduction and feature extraction because it is an effective deep feature extractor. High-level spatial and contextual features in cervical images were learnt automatically, through the convolutional backbone of YOLOv12. The emphasis on salient areas and the inhibition of unnecessary background details is useful as it can reduce the dimensionality of the feature representations and maintain discriminative properties of the version of YOLOv12. Deep features extracted further underwent refinement with the help of an Ensemble Boosting Algorithm (EBA) taking into consideration the informative features and discarding irrelevant or redundant ones. The concurrence of all these strategies makes the prediction model computationally simpler, thus avoiding overfitting[19] and improving the generalization ability of the model. Comprehensively, the chosen pre-processing image and feature extraction plan based on YOLOv12 can guarantee the quality of images, less noise, and features, which are excellent in the precision, efficacy, and consistency of prediction models of cervical cancer. Figure 5 shows the images after preprocessing.

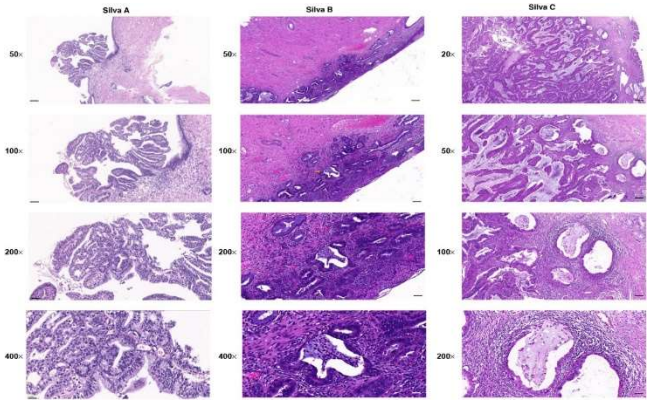

Figure 5: Shows the Images after Preprocessing.

4.2 Cervical Cancer Prognostication Two-Phase Proposal.

In the proposed structure, the prediction of cervical cancer is performed in two separate stages that will guarantee effective learning of features and correct classification. The general process of work is in compliance with general machine learning evaluation procedures, such as data collection, preprocessing, training, testing, and performance visualization[20].

4.3 Data Pre-processing and Data Collection.

The data on cervical cancer is gathered in the initial step, which involves various repositories, which are publicly available. The obtained samples are pre-treated to deal with missing values, normalize the range of features, and eliminate noise. All the samples are classified as either positive (+ve) or negative (-ve) with respect to the occurrence or absence of cervical cancer. The processed data is further split into training and testing set to determine the generalization ability of the model proposed.

Phase 1: Phase 1 Utilizes the YOLOv12 to extract features.

Phase 1 is when the already processed cervical cancer dataset is inputted into the YOLOv12 architecture

which is used as a sophisticated feature extractor. YOLOv12 effectively trains in discriminative spatial and contextual attributes of the input data, especially clinically meaningful attributes of the data, including nuclear chromatin properties, cell morphology, and changes in texture.

YOLOv12 network works with the input samples with several convolutional layers to generate high-dimensional feature vectors that can capture local and global information to be used in cervical cancer detection. Such obtained deep features create a polished image of the dataset and are used as input in the classification phase.

Phase 2: RF -XGBoost Ensemble Classification.

Phase 2: YOLOv12 deep features are then fed into a hybrid Random Forest (RF) and XGBoost ensemble classifier (RF -XGBoost) to perform the final classification. First, the Random Forest model was used to minimize the variance of features and enhance stability by combining several decision trees. The XGBoost is then used to further improve the accuracy of the prediction by using gradient boosting to reduce the error of the classification. The ensemble classifier is trained with the help of the training feature set and the labels:

Z = RF_XGBoost_Classifier()

Z.fit(X_train, Y_train)

Tests are performed on the training model after training:

Y_pred = Z.predict(X_test)

Calculation of the confusion matrix is done to evaluate the classification performance using the true and forecasted labels. The precision of the suggested model is calculated as: The RF -XGBoost classifier using ensemble shows better accuracy and stability over the conventional single classifiers. This two-stage YOLO v12 RF XGBoost[21] architecture is a powerful combination of deep feature representation and ensemble learning, which results in effective and high-quality cervical cancer prediction. The entire cycle of the proposed model is presented in Figure 6.

4.4 Algorithm for proposed classifier:

Input:
Pre-processed Cervical Cancer Dataset

Output:
Predicted Class Labels, Confusion Matrix, and Accuracy

Phase I: Feature Extraction Using YOLOv12

Step 1:
Collect cervical cancer data from publicly available repositories.

Step 2:
Pre-process the collected data by performing noise removal, normalization, and handling missing values.

Step 3:
Label each sample as **positive (+ve)** or **negative (– ve)** based on cervical cancer diagnosis.

Step 4:
Split the dataset into training and testing sets:

- Training set: X_{train}, Y_{train}
- Testing set: X_{test}, Y_{test}

Step 5:
Initialize the YOLOv12 model for deep feature extraction.

Step 6:
Feed the training samples into the YOLOv12 network.

Step 7:
Extract deep features representing cervical cancer indicators such as nuclear chromatin distribution, cell morphology, and texture patterns.

Step 8:
Store the extracted features as structured feature vectors for the classification phase.

Phase II: Classification Using RF–XGBoost Ensemble

Step 9:
Initialize the hybrid ensemble classifier:

$$Z = \text{RF–XGBoost Classifier}()$$

Step 10:
Train the classifier using the extracted feature vectors and corresponding labels:

$$Z.\text{fit}(X_{train}, Y_{train})$$

Step 11:
Predict class labels for the test dataset:

$$Y_{pred} = Z.\text{predict}(X_{test})$$

Step 12:
Generate the confusion matrix using actual and predicted labels:

$$P = \text{ConfusionMatrix}(Y_{test}, Y_{pred})$$

Step 13:
Compute the classification accuracy:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$$

Step 14:
Visualize the prediction results and performance metrics.

Step 15:
Return the confusion matrix and accuracy as the final output.

Stop

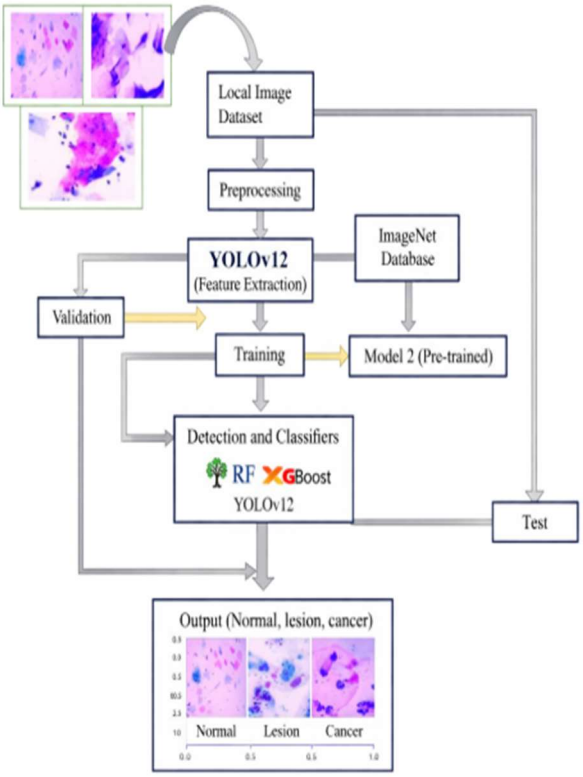

Figure 6: Shows How Proposed Classifier Works In Predicting Cervical Cancer

4.5 Support Vector Machine(SVM)

The goal of the model in cervical cancer prediction is to identify the intricate patterns in high dimensional clinical and diagnostic measurements in order to adequately differentiate between the cases of cancers and non-cancerous cases. One of the most effective supervised machine learning methods applicable towards this end is the Support Vector Machine[22] (SVM) because it has a strong generalization performance and its ability to work with both linear and non-linear data. SVM model seeks to find an optimal point in the multidimensional feature space, by building a decision boundary, referred to as a hyperplane, that optimally classifies various classes. In a three-dimensional (or higher dimensional) space, there can be several hyperplanes which are able to partition the data. Nonetheless, SVM specifically chooses the hyperplane that has the largest margin (largest distance between the hyperplane and the closest data point in each class (so-called support vectors). This margin can be maximized to enhance the effectiveness of the model in categorizing cases of unseen cervical cancer[23] with greater certainty and

lower chances of overfitting. When the optimum hyperplane has been determined, the data of future patients (age, HPV status, cytology results, biopsy results and clinical symptoms) can be correctly categorized in benign or malignant groups .To compute this hyperplane, SVM method builds this hyperplane in a fairly high-dimensional or even an infinite-dimensional space of mathematical transformations, which gives it the opportunity to successfully address complicated associations among features. In addition to the classification aspect, SVM can also be used in regression, feature extraction and noise filtering, and therefore, it can be used in medical prediction systems where the variability of data is high. Two SVMs apply in predicting cervical cancer namely Linear SVM and Non-Linear SVM. Linear SVM is used in the case the data of cervical cancer are linearly separable, i.e. two classes can be separated by a straight line (or plane in a higher dimension). Linear SVM classifier[24] is effective in such cases where the single decision boundary of cancer-positive and cancer-negative samples can be identified. This method is computationally efficient, and it is good when features have definite linear relationships. Nevertheless, medical data that can be observed in real life, such as cervical cancer datasets, cannot be separated non-linearly since they are usually characterized by overlapping patterns and interactions among risk factors. In such cases, Non-Linear SVM is used. This classifier works with the use of kernel functions, including radial basis function (RBF) or polymer kernel to transform original data into a higher dimensional space where separation can occur on a linear basis. Non-Linear SVMs, therefore, contribute to a large improvement in the accuracy of prediction of cervical cancer by recognizing complex dependencies of features. Altogether, SVM is a valid and effective tool to predict cervical cancer and enhance clinical decision-making, which can lead to early detection and improve clinical outcomes. Figure 7 shows data classification using linear svm. And in Figure 8 data classified using non-linear svm.

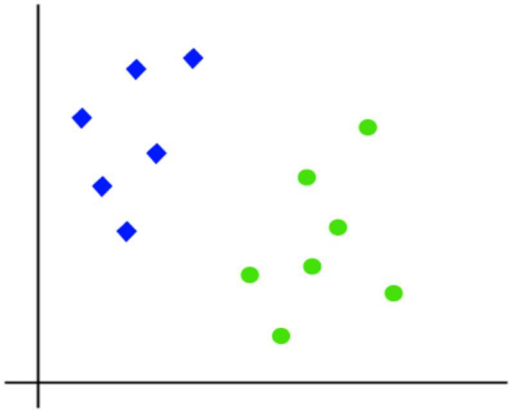

Figure 7: Shows Data Is Classified Into Labels In 1-D.

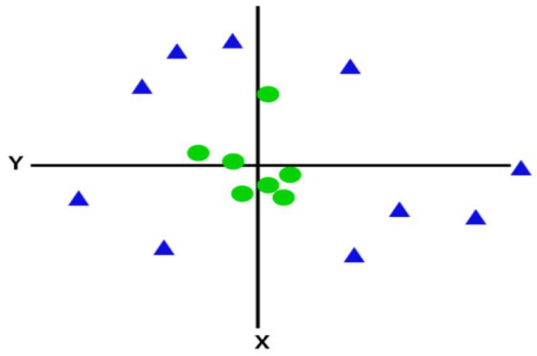

Figure 8: Shows Data Is Classified Into Labels In 1-D Of Non-Linear SVM.

4.6 Decision Tree:

Decision Tree (DT) algorithm[25] can be utilized to resolve problems involving classification and regression. When applied to cervical cancer prediction, decision trees are especially handy since they offer interpretable decision-making paths which can guide the clinician to get a better idea of the way in which predictions are achieved. The algorithm is named so because it is structurally similar to a tree, the branches of which are decision rules and the nodes of which are the outcomes. The procedure starts at the root node which is the entire dataset of cervical cancer and moves to decision nodes whereby the data is divided according to a particular condition of features and finally reaches the leaf nodes which give the outcome of the predicted classification. To explain such process, a decision tree built based on a cervical cancer prediction dataset can be considered. The training data is first broken at the root node by considering all the available attributes as age, HPV

status, number of pregnancies, smoking habits and cytology results. Assuming that there are three features, the algorithm comes up with three candidate splits[26]. To measure the quality of separation of the data, the cost or impurity function is computed using each possible split. The split which limits the impurity (or cost) is the one chosen as the best split. This recursive process is done on the next set created and so on making the tree to grow exponentially. The decision tree algorithm is considered a greedy algorithm because it is a step-by-step selection of the optimum split at every node, and the algorithm never reevaluates past choices. To classify the data, the Gini index is often employed to define the quality of a split[27], and it is defined as:

$$C=\sum pk(1-pk)$$

where p_krepresents the proportion of samples belonging to class kwithin a node. The Gini index indicates the extent of impurity of classes in the resulting subsets. A node that has a balanced number of cancer-positive and cancer-negative samples (i.e.,<|human|>A node with an equal number of cancer-positive and cancer-negative samples (i.e., pk=0.5). The impurity is greatest at =0.5) which gives a Gini number of 0.5 in binary classification. A node on the other hand is perfectly pure when every sample is of one type and the Gini index is 0. This level of impurity is used to direct the splitting to enhance predictive results. Figure 6 gives an example of a decision tree created on the cervical cancer prediction data. Developing on this, the Decision Tree Algorithms has two phases of operation. During the initial step, a decision tree is built based on cervical cancer data retrieved on the Kaggle repository. The second phase has an ensemble boosting mechanism implemented on the base decision tree to improve its classification performance. The enhanced version assesses the performance based on a confusion matrix and is more accurate than the existing classifiers. Each node in the recursive expansion of the tree is chosen with the help of the attribute which makes the greatest contribution to the separation of the classes. When the splitting does not increase the accuracy of prediction and when all samples in a subset are of the same type, the process is terminated. Figure 9 illustrates the entire process of the DT model.

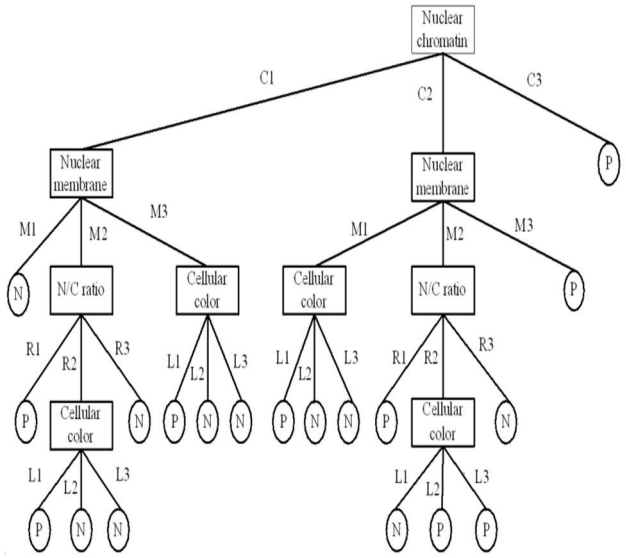

Figure 9 Shows Construction Of DT For Predicting CC.

In the 1st phase based on the dataset taken from Kaggle repository a Decision tree is constructed. Each edge to a node's children corresponds to a different input variable, and the goal of the model is to predict the value of the target variable. As the tree grows recursively based on the information gained, the attribute with the greatest information gain is selected. The process will be complete when the value of the subset coincides with the value of the target variable or when splitting no longer improves the prediction[28]. Figure 10 shows the growth level of cervical cancer.

To find information gain for the given dataset entropy of D is defined as:

$$Entropy(D) = -\sum_{i=1}^{m} p_i \log_2(p_i)$$

So, from the entropy we find the information gain for the cervical data which is given as:

$$Gain(D, A) = Entropy(D) - \sum_{j=1}^{v} \frac{|D_v|}{|D|} Entropy(D_v)$$

The attribute with best normalized information gain is chosen for the best split which is given as:

$$Split(D,A)=\sum_{j=1}^{v}\frac{|D_v|}{|D|}\times\log_2\left(\frac{|D_v|}{|D|}\right)$$

From the split and gain we found the gain ratio which plays a major role for constructing Decision tree given as:

$$GainRatio(D,A)=\frac{Gain(D,A)}{Split(D,A)}$$

In the Figure 8 it shows the transformation of cervical cancer from normal stage to advance.

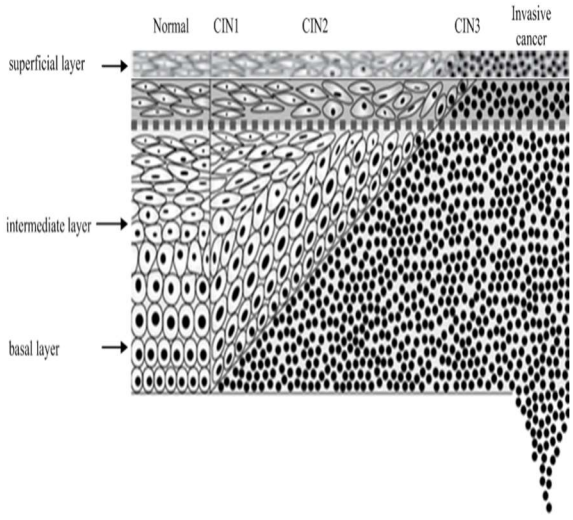

Figure 10: Shows Cervical Cancer From Normal To Massive Level

5. IMPLEMENTATION & RESULTS

There are specific guidelines that must be followed for any research problem in order to predict the accuracy, such as sensitivity and specificity. F-score. Sensitivity typically indicates whether or not a test is capable of correctly identifying certain conditions, such as precancer. The accuracy of the classifier may suffer because the higher the sensitivity, the fewer precancer cases we can detect. The ability to correctly identify precancer cases is tested using the second measure of specificity. The higher the specificity, the fewer cases of normal cervix we can expect, which may result in model inaccuracy. Therefore, maintaining both in balance may lead to the best forecasting accuracy. Generally, Specificity, Sensitivity can be calculated based on 2 parameters that are obtained from confusion matrix which is shown in Figure 11.

$$Accuracy=(TP+TN)/(TP+TN+FP+FN) \quad (1)$$

$$Precision=TP/FP+TP \quad (2)$$

$$Recall=TP/FN+TP \quad (3)$$

$$F1=2(Precision * Recall)/(Precision + Recall) \quad (4)$$

		Real	
		Positive	Negative
Predicted	Positive	True Positive (tp)	False Positive (fp)
	Negative	False Negative (fn)	True Negative (tn)

Figure 11: Shows The Generalized Confusion Matrix To Find Specificity, Sensitivity.

The following Figure 12. Shows the count of people who has risk of getting cervical cancer based on their ages and also Figure 13 shows count of sexual partners who has a risk of getting cancer. And Figure 14 tells the count based on First sexual intercourse. The newly proposed YOLOv12 and RF-XGBoost framework for the prediction of cervical cancer has a wide range of practical applications in the healthcare industry including automated medical diagnosis and intelligent clinical decision support systems. The model can help doctors detect cervical cancer early in the diagnosis process by giving medical images accurate and reliable risk predictions and reducing diagnostic errors. This is particularly helpful in the developing world where trained oncologists and screening centers are scarce. By using deep feature extraction and ensemble learning, this method can facilitate the rapid and stable diagnosis of the disease, which is usually performed manually. The proposed system can be applied in hospitals, diagnostic laboratories, telemedicine platforms and AI-driven healthcare applications to enhance the efficiency of screening, reduce the workload of clinical practitioners and reduce the inter-observer error. Additionally, its adaptability to both diverse data sources and class imbalances make it well-suited for large-scale real-world applications in modern healthcare sectors such as precision medicine and AI-driven diagnostics.

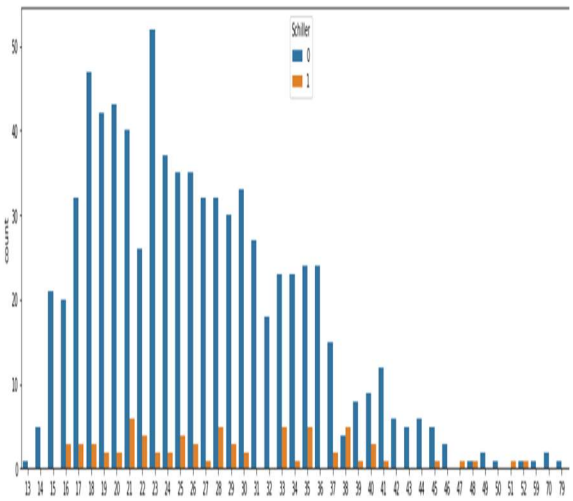

Figure 12. Show The Count Vs Age

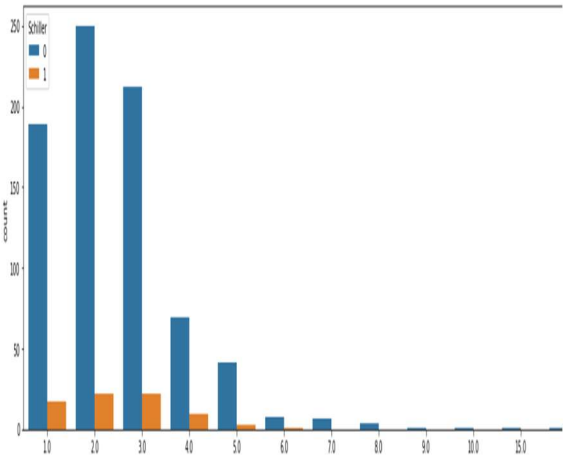

Figure 13. Show The Count Vs No of Sexual Partners

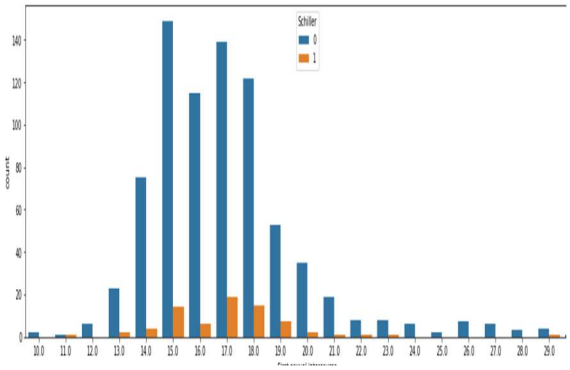

Figure 14. Show The Count Vs First Sexual Intercourse

A heat map is generated for predicting the accuracy using the following testing parameters like sensitivity, specificity and accuracy of the proposed classifier is shown in Figure 15. Table 2 shows the confusion matrix obtained by proposed classifier. And Figure 15 shows the performance metrics obtained by proposed classifier. And Table 3 shows the validation table obtained by proposed classifier. Table 4 shows the confusion matrix obtained by DT classifier. And Figure 16 shows the performance metrics obtained by DT classifier. And Table 5 shows the validation table obtained by DT classifier. Table 6 shows the confusion matrix obtained by DT classifier. And Figure 17 shows the performance metrics obtained by DT classifier. And Table 7 shows the validation table obtained by DT classifier.

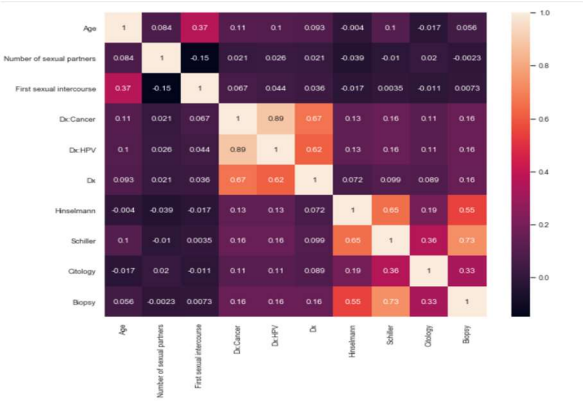

Figure 15: Shows Generated Heat Map

Table 2: Confusion Matrix Obtained for Proposed Classifier

Predicted Class \ Actual Class		
	22972	724
	876	14128

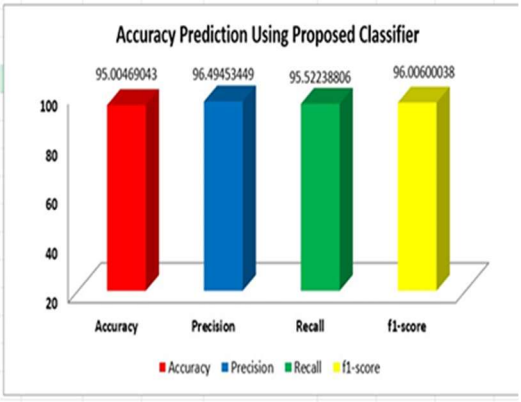

Figure 15: Shows the Performance Metrics for Proposed Classifier.

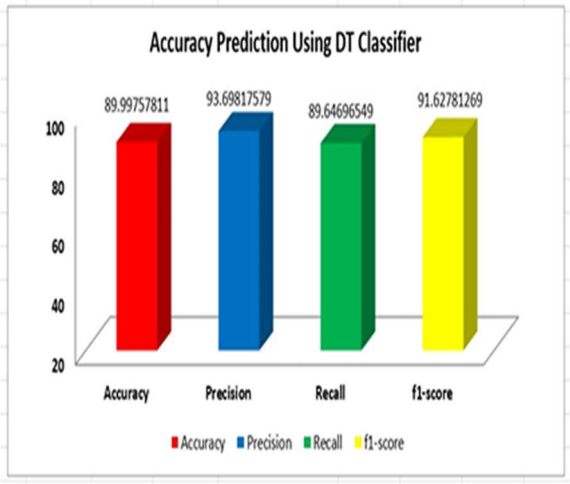

Figure 16: Shows the Performance Metrics for DT Classifier.

Table 3: Accuracy Table for Proposed Classifier

Shows the accuracy of Proposed Algorithm

Label	Precision	Recall	F-Score	Support
0	91.13	96.15	93.29	21696
1	92.15	95.28	92.19	17004
Accuracy			95.97	38700
MicroAvg	91.83	96.05	93.09	38700
MicroAvg	92.15	96.98	93.59	38700

Table 4: Confusion Matrix Obtained for DT Classifier.

Predicted Class \ Actual Class		
	20670	1790
	2012	14228

*Table 5: Accuracy Table for **DT** Classifier*

Shows the accuracy of **DT** Algorithm

Label	Precision	Recall	F-Score	Support
0	90.23	89.19	89.70	21460
1	89.85	91.98	90.59	17240
Accuracy			90.95	38700
MicroAvg	90.23	89.19	89.70	38700
MicroAvg	89.85	91.98	90.59	38700

Table 6: Confusion Matrix Obtained for SVM Classifier.

Predicted Class \ Actual Class		
	20265	2560
	2342	13533

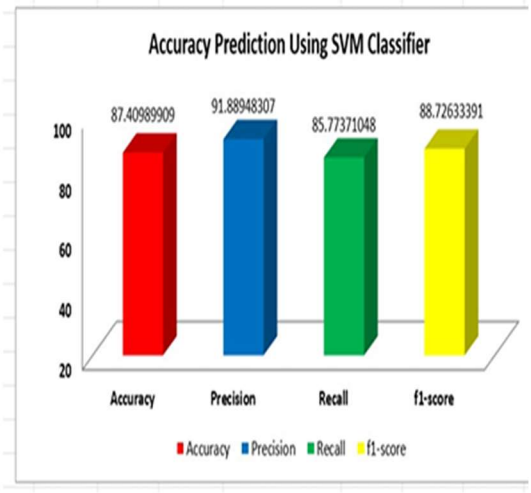

Figure 17: Show the Performance Metrics for SVM Classifier.

*Table 7: Accuracy Table for **SVM** Classifier*

Shows the accuracy of **SVM** Algorithm

Label	Precision	Recall	F-Score	Support
0	90.81	88.75	89.79	21025
1	90.15	88.98	89.59	17675
Accuracy			87.95	38700
MicroAvg	90.81	88.75	89.79	38700
MicroAvg	90.15	88.98	89.59	38700

6. CONCLUSION AND FUTURE WORK:

This research aimed to develop a hybrid machine learning approach that combines YOLOv12 deep feature extraction with RF-XGBoost ensemble classification to analyze the risk of cervical cancer. The developed model has been able to enhance the prediction and classification of cervical cancer with the help of significant physiological and clinical parameters. Through experimental evaluation, the accuracy, precision, recall, and F1 score of the proposed hybrid framework were compared with those of traditional machine learning classifiers and it was shown that the proposed framework has higher predictive accuracy, robustness and generalization capability. YOLOv12 demonstrated excellent automatic deep feature extraction capabilities, capturing discriminative patterns, and RF-XGBoost ensemble improved the stability of classification and decreased the errors in diagnosis. The proposed approach was compared with conventional statistical and machine learning methods, and its superiority was proved by comparative analysis. The results obtained show that the proposed system is an efficient and reliable medical diagnosis support system that can be used for identification of a person at an early stage of cervical cancer and for automatic diagnosis of the disease. Future research can extend the proposed YOLOv12 and RF-XGBoost framework, such as the inclusion of larger and more varied datasets of cervical cancer for model generalization and robustness. Prediction accuracy and risk stratification could be further enhanced by the integration of multimodal medical data such as histopathological images, clinical records, and genomic information. Optimizing features extraction and computational complexity can also be achieved through exploring advanced deep learning architectures and attention-based mechanisms. Moreover, explainable AI methods can be incorporated to enhance the interpretability and trust of AI-driven predictions by clinicians. The proposed framework can also be studied for real-time deployment in cloud-based and telemedicine healthcare systems for large-scale clinical applications in the future.

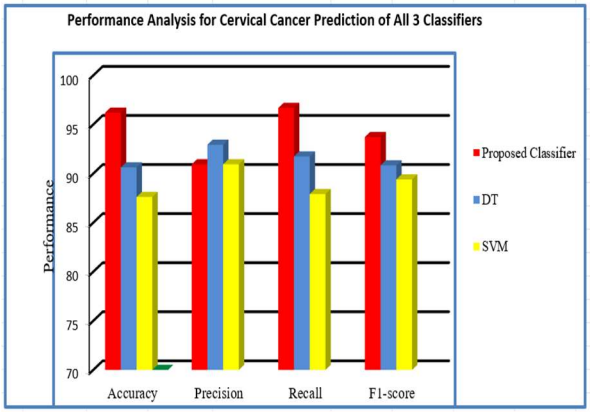

Figure 18: Shows the Performance of All 3 Classifiers.

REFERENCES

[1]. Xie R, Sha S, Brenner H, Schöttker B. Circulating inflammation-related proteome improves cardiovascular risk prediction. Results from two large European cohort studies. Eur J Epidemiol. 2025;40(10):1191–203

[2]. Qi X, Wang S, Fang C, Jia J, Lin L, Yuan T. Machine learning and SHAP value interpretation for predicting comorbidity of cardiovascular disease and cancer with dietary antioxidants. Redox Biol. 2025;79:103470.

[3]. Oexner RR, Ahn H, Theofilatos K, Shah RA, Schmitt R, Chowienczyk P, Zoccarato A, Shah AM. Serum metabolomics improves risk stratification for incident heart failure. Eur J Heart Fail. 2024;26(4):829–40.

[4]. T. V. Sai Krishna, S. K. Pathuri, K. S. Sandeep, M. K. Padhi, I. Aswani and D. Haritha, "An Efficient Machine Learning Classification Model for Diabetes Prediction," 2023 IEEE 12th International Conference on Communication Systems and Network Technologies (CSNT), Bhopal, India, 2023, pp. 139-146, doi: 10.1109/CSNT57126.2023.10134615.

[5]. Pathuri, S.K.; Anbazhagan, N.; Joshi, G.P.; You, J. Feature-Based Sentimental Analysis on Public Attention towards COVID-19 Using CUDA-SADBM Classification Model. Sensors 2022, 22, 80. https://doi.org/10.3390/s22010080.

[6]. S. K. Pathuri, N. Anbazhagan and G. B. Prakash, "Feature Based Sentimental Analysis for Prediction of Mobile Reviews Using Hybrid Bag-Boost algorithm," 2020 7th International Conference on Smart Structures and Systems (ICSSS), Chennai, India, 2020, pp. 1-5, doi: 10.1109/ICSSS49621.2020.9201990.

[7]. Pathuri SK, Anbazhagan N, Joshi GP, You J. Feature-Based Sentimental Analysis on Public Attention towards COVID-19 Using CUDA-SADBM Classification Model. Sensors. 2022; 22(1):80. https://doi.org/10.3390/s22010080.

[8]. A. Gabriel et al., "Mask-R-CNN for sugarcane weed identification using ResNet-50 and ResNet-101 backbones," Agriculture, 2022–2023.

[9]. S. K. Pathuri, N. Anbazhagan, M. K, T. V. S. Krishna and C. Prasad, "A Hybrid Bag-Boost algorithm for Feature-Based Sentimental Analysis on Mobile Reviews," 2021 2nd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2021, pp. 948-953, doi: 10.1109/ICOSEC51865.2021.9591947.

[10]. W. Luo, "Predicting cervical cancer outcomes: statistics, images, and machine learning," Frontiers in Artificial Intelligence, vol. 4, article 627369, 2021.

[11]. A. Jajodia, A. Gupta, H. Prosch et al., "Combination of radiomics and machine learning with diffusion-weighted MR imaging for clinical outcome prognostication in cervical cancer," Tomography, vol. 7, no. 3, pp. 344–357, 2021.

[12]. D. Ding, T. Lang, D. Zou et al., "Machine learning-based prediction of survival prognosis in cervical cancer," BMC Bioinformatics, vol. 22, no. 1, pp. 331–331, 2021.

[13]. L. Akter, M. M. Ferdib-Al-Islam, M. S. Islam, M. R. Al-Rakhami, and Haque, "Prediction of cervical cancer from behavior risk using machine learning techniques," SN Computer Science, vol. 2, no. 3, 2021.

[14]. A. Arora, Tripathi, and Bhan, "Classification of cervical cancer detection using machine learning algorithms," in 2021 6th International Conference on Inventive Computation Technologies (ICICT), pp. 1075–1098, Coimbatore, India, 2021.

[15]. R. Weegar and K. Sundström, "Using machine learning for predicting cervical cancer from Swedish electronic health records by mining hierarchical representations," PLoS One, vol. 15, no. 8, pp. 237911–237911, 2020.

[16]. T. R. Ramesh, U. K. Lilhore, M. Poongodi, S. Simaiya, A. Kaur, and M. Hamdi, "Preditive analysis of heart diseases with machine learning approaches," Malaysian Journal of Computer Science, pp. 132–148, 2022.

[17]. S. Kim, S. Lee, C. H. Choi et al., "Machine learning models to predict survival outcomes according to the surgical approach of primary radical hysterectomy in patients with early cervical cancer," Cancers, vol. 13, no. 15, p. 3709, 2021.

[18]. D. T. S. Patel, "A cross sectional study to estimate delay in diagnosis and treatment of tuberculosis (TB) among patients attending urban health centre in an urban slum area," Public Health Review: International Journal of Public Health Research, vol. 5, no. 1, pp. 1–7, 2018.

[19]. L. Gupta, A. Edelen, N. Neveu, A. Mishra, C. Mayes, and Y. K. Kim, "Improving surrogate model accuracy for the LCLS-II injector

frontend using convolutional neural networks and transfer learning," Machine Learning: Science and Technology, vol. 2, no. 4, pp. 1245–1265, 2021.

[20]. [Kondratenko Y., Atamanyuk I., Sidenko I., Kondratenko G., Sichevskyi S. Machine Learning Techniques for Increasing Efficiency of the Robot's Sensor and Control Information Processing. Sensors. 2022;22:1062.,doi: 10.3390/s22031062.

[21]. [Hall M.A. Correlation-Based Feature Selection for Machine Learning. The University of Waikato; Hamilton, New Zealand: 1999.

[22]. Dokduang K., Chiewchanwattana S., Sunat K., Tangvoraphonkchai V. A comparative machine learning algorithm to predict the bone metastasis cervical cancer with imbalance data problem. Recent Adv. Inf. Commun. Technol. 2014;10:93–102.

[23]. Ghoneim A., Muhammad G., Hossain M.S. Cervical cancer classification using convolutional neural networks and extreme learning machines. Future Gener. Comput. Syst. 2020;102:643–649. doi: 10.1016/j.future.2019.09.015.

[24]. Ashok B., Aruna P. Comparison of Feature selection methods for diagnosis of cervical cancer using SVM classifier. Int. J. Eng. Res. 2016;6:94–99.

[25]. Janbhasha, Shaik, Venkateswarlu Sunkari, Dipanwita Debnath, Venkat Rao Pasupuleti, Aketi Naresh, Lakshmanarao Talapakula, And Jutu Gopaiah. "Integrating Adversarial Autoencoders with Gated Recurrent Units to Identify Anomalies in Ecg Signals." Journal Of Theoretical and Applied Information Technology 104, No. 5 (2026).

[26]. S. Janbhasha, C. H. N. S. Kumar, V. Sitharamulu, B. N. V. Madhu Babu, H. R. Battu, and K. Venkataramana, "A Hybrid Approach for Fraud Detection in Digital Wallet Transactions Using Adversarial Autoencoders and Gated Recurrent Units", Eng. Technol. Appl. Sci. Res., vol. 15, no. 4, pp. 25532–25537, Aug. 2025.

[27]. Benti, D.M., Janbhasha, S., Desisa, E.G. (2023). Identification of Generative Adversarial Network Forms, Open Issues, and Future Study Areas: A Study. In: Fong, S., Dey, N., Joshi, A. (eds) ICT Analysis and Applications. Lecture Notes in Networks and Systems, vol 517. Springer, Singapore. https://doi.org/10.1007/978-981-19-5224-1_31.

[28]. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians. 2021; 71: 209–249