SN Computer Science

Prediction of Cervical Cancer with Machine Learning Approaches

--Manuscript Draft--

Manuscript Number:	
Full Title:	Prediction of Cervical Cancer with Machine Learning Approaches
Article Type:	Original Research
Section/Category:	Emerging Trends in Sensors, IoT and Smart Systems
Funding Information:	
Abstract:	Cervical cancer is the second most common disease among Indian women aged 15 to 44, is caused by aberrant cell proliferation in the cervix. So, its early detection very crucial. Many screening measures like as Pap smears, HPV tests and Colposcopies are supported in this scenario. Inception + Support Vector Machine (SVM), Ensemble method,), K-Nearest Neighbour (KNN), Bagging Decision Tree (DT), Logistic Regression (LR), Convolutional Neural Networks (CNNs) Machine Learning techniques are the fundamental blocks of the system for cervical cancer cell identification and classification presented in this paper. In addition to the normal classification algorithms used to identify cervical cancer, cell segmentation, and feature extraction methods are typically required. Also, these proposed models need a huge dataset to avoid overfitting and poor in generalization problems. After integrating the cell images into these models to obtain deep-learning features, an Extreme Learning Machine (ELM)-based classifier classifies the given or input images. These techniques are also used for transfer learning and fine-tuning. These proposed models use the Inception deep learning model, which diagnoses cervical cancer in multiple classes with an precision of approximately 99.9% and gives accuracies based on the input dataset.
Corresponding Author:	Neeru Jindal Thapar Institute of Engineering and Technology: Thapar Institute of Engineering and Technology (Deemed to be University) INDIA
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Thapar Institute of Engineering and Technology: Thapar Institute of Engineering and Technology (Deemed to be University)
Corresponding Author's Secondary Institution:	
First Author:	Aditya Pandey
First Author Secondary Information:	
Order of Authors:	Aditya Pandey
	Jatin Arora
	Gursamrath Singh R. Kohli
	Mehar Khurana
	Neha Sharma
	Neeru Jindal
Order of Authors Secondary Information:	
Author Comments:	Dear Sir I am submitting manuscript for consideration in your esteemed journal. Regards

Suggested Reviewers:	Dr. Kanwarpreet, Ph D kpreet2392@gmail.com She is potential Reviewer
	Dr. Manoj Kumar manojsingla77@gmail.com He is a potential Reviewer.

Prediction of Cervical Cancer with Machine Learning Approaches

Aditya Pandey[1*], Jatin Arora[1], Gursamrath Singh R. Kohli[1], Mehar Khurana[1], Neha Sharma[2], Neeru Jindal[1**]

[1]*Department of Electronics and Communication Engineering, Thapar Institute of Engineering and Technology, Patiala (Punjab), India.*
[2]*Department of Computer Science and Engineering, Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India.*

apandey2_be22@thapar.edu, jarora2_be22@thapar.edu, , gkohli_be22@thapar.edu, mkhurana_be22@thapar.edu, neha.1129@chitkara.edu.in, neeru.jindal@thapar.edu

Aditya Pandey*-First Author.
Neeru Jindal**- Corresponding Author.

Abstract
Cervical cancer is the second most common disease among Indian women aged 15 to 44 and is caused by aberrant cell proliferation in the cervix. So, its early detection is very crucial. Many screening measures, such as Pap smears, HPV tests, and Colposcopies, are supported in this scenario. Inception + Support Vector Machine (SVM), Ensemble method,), K-Nearest Neighbour (KNN), Bagging Decision Tree (DT), Logistic Regression (LR), Convolutional Neural Networks (CNNs) Machine Learning techniques are the fundamental blocks of the system for cervical cancer cell identification and classification presented in this paper. In addition to standard classification algorithms used to identify cervical cancer, cell segmentation and feature extraction methods are typically required. Also, these proposed models have massive datasets to avoid overfitting or generalization problems after integrating the cell images into these models to obtain deep-learning features. An Extreme Learning Machine (EL-based classifier classifies the given or input images. These techniques are also used for transfer learning and fine-tuning. These proposed models use the Inception deep learning model, which diagnoses cervical cancer in multiple classes with a precision of approximately 99.9% and gives accuracies based on the input dataset.

Keywords Machine learning. Deep learning. Cervical cancer. Transfer learning. Classification

Introduction

The tumour that can grow near the base of the womb is referred to as cervical cancer [1]. It happens when the cells change in the cervix that connects the uterus and the vagina of the female reproductive organ and slowly spreads to the rest of the body parts.

Cervical cancer is generally diagnosed in females belonging to the age group of 35 to 44, and over 20% of the cases are diagnosed in females after the age of 65. In 2020, around 6 lakh new cases were reported globally, and around 3.5 lakh deaths were caused, making it the fourth most common cancer globally [2-4]. The human papillomavirus, or HPV, is a significant factor in the spread of cervical cancer during intercourse. Generally speaking, HPV is the reason penile cancers (60%), anal and cervical cancers (90%), and roughly vaginal and vulvar cancers (70%). When the female body's immune system does not eliminate an HPV infection comprising oncogenic HPV types, the infection may linger and eventually lead to the development of malignant cells from abnormal ones. Cervical HPV infections enhance the risk of developing cervical cancer, with 10% of infected women going on to have persistent infections. Early on, a precancerous lesion of the cervix frequently shows no symptoms or indicators. However, these cells eventually disperse throughout the body, so the

First Author: Aditya Pandey[1]*- apandey2_be22@thapar.edu
Corresponding Author: Neeru Jindal[1]** - neeru.jindal@thapar.edu
Jatin Arora[1]- jarora2_be22@thapar.edu, Gursamrath Singh R. Kohli[1] - gkohli_be22@thapar.edu, Mehar Khurana[1]- mkhurana_be22@thapar.edu, Neha Sharma[2] - neha.1129@chitkara.edu.in,

[1]*Department of Electronics and Communication Engineering, Thapar Institute of Engineering and Technology, Patiala (Punjab), India.*
[2]*Department of Computer Science and Engineering, Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India.*

symptoms become apparent. The degree to which these symptoms are unique will vary based on the tissues or organs where it has progressed. Early detection of cervical cancer is indicated by light bleeding or blood spots during or later periods. It is longer and heavier than usual menstrual flow; bleeding occurs after sexual activity, pelvic exam or douching, increased vaginal discharge, pain through sexual activity, and bleeding after menopause. These symptoms are identified using various technological advancements in the medical fields, including CT scans (Computed tomography), MRI (Magnetic Resonance Imaging), and Pap Smear Tests that detect cell changes in the cervix region of the reproductive system. The cervix thin, the lower end of the uterus that is placed at the top of the vagina, is where cells are extracted for a Pap smear.

It might also help identify other illnesses like inflammation or infections. HPV DNA tests are an additional technique that is utilized to identify cervix cells that may be susceptible to HPV-related malignancy. A complete evaluation of your cervix is probably the first step in the cervical cancer testing process. A colposcopy is a specialized magnifying device used to look for cancerous growths. A sample of cervical cells is taken for laboratory analysis during the colposcopic examination. Fig.1 presents the flow work for cervical cancer and its class prediction.

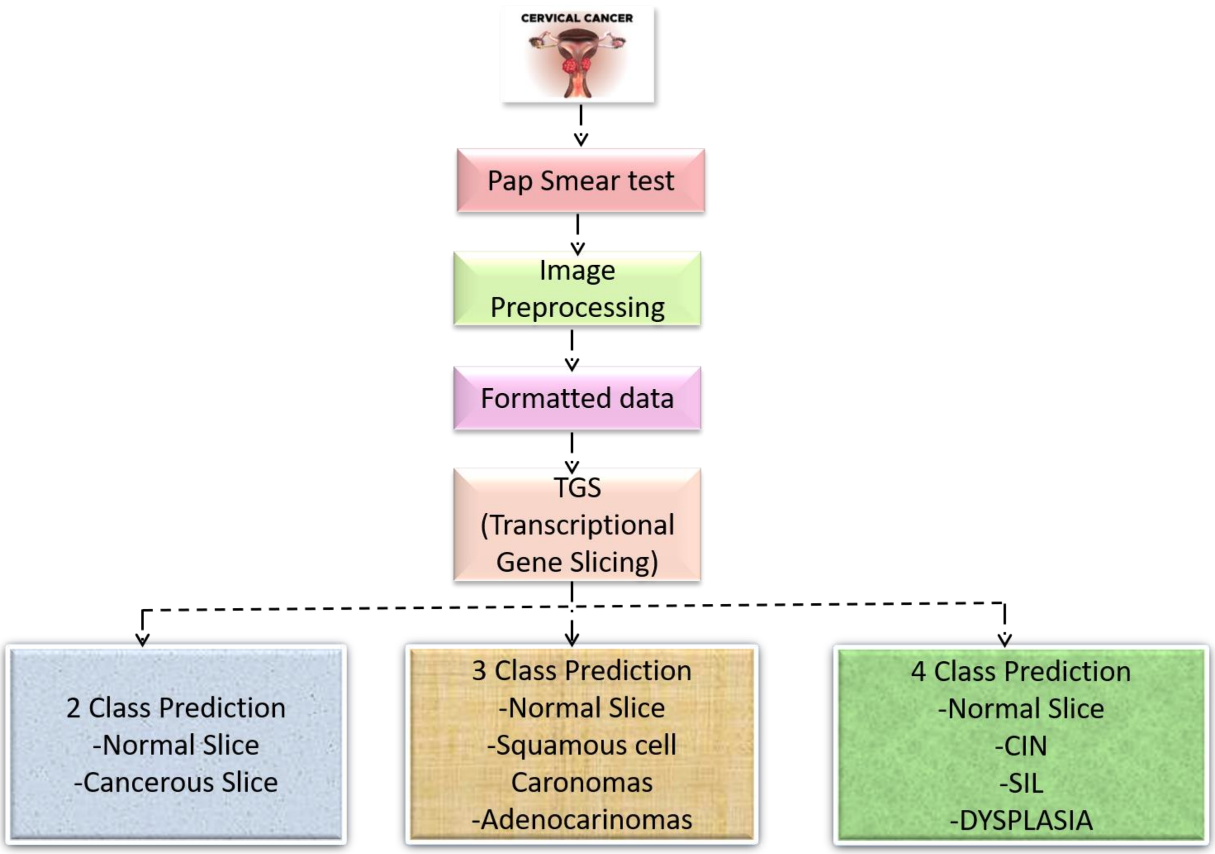


Fig. 1 The process for cervical cancer and its class prediction.

These samples are further examined and processed into the images, which are sent to the examination center for further diagnosis. These images are preprocessed with the help of various ML and DL models, and predictions are made using them. Information from the real world is often partial and erroneous, for example. In keeping with this, raw data can be accurately represented and applied to the dataset by changing and cleaning it to facilitate a reliable analytical delivery. The dataset on cervical cancer obtained for analysis has noise, missing values, and redundancies. In light of the growing significance of health-related concerns, mining methods are regarded as one of the most critical and challenging areas of medical research. By utilizing the insights it acquires from training machine learning models on various cervical cancer datasets, the data mining methods can improve the screening procedure for the disease. The medical community uses these techniques to find trends and linkages between these symptoms and cancer prognosis.

Numerous data mining approaches can be used to quickly recommend further research and medical care, which can save lives, particularly in the case of cervical cancer. Preprocessing 80 percent of the data is the initial phase and is an essential part of any data mining operation.

Our research's key components that raise the level of the training dataset are identified by the machine learning models. Based on the sample data provided, it results in the classification of cervical cancer into multiple classes. Machine learning is ideal for analyzing specific biological data in pharmacogenomics research findings since it offers several noteworthy advantages. Furthermore, a large variety of qualitative and quantitative input vectors can be accommodated by it. Secondly, it evaluates the property's significance in determining the type, offering a standard for choosing features. SVM, InceptionV3, SVM, Bagging DT, Logistic Regression, KNN, CNN, and many other ML and DL techniques are applied to provide adequate accuracy for predicting cervical cancer. This work classifies cervical cells as malignant or noncancerous using a direct method to identify cervical carcinomas using public databases, enabling simple comparison and accessibility for the research community [5].

The main contributions of the research paper are:

1. The proposed work takes a direct approach by automatically detecting cervical carcinomas from the public dataset by processing Pap smear images directly. To make it easier for the research community to compare and access the models and results, which were compiled under similar experimental conditions, it was proposed that cervical cells be classified as either cancerous or noncancerous and that the type of cervical cancer be further identified.
2. The suggested approach provides a great generalization to the model by applying data augmentation techniques to increase the data.
3. Various max-pooling and convolutional layers have been applied to get good prediction accuracy.
4. This proposed work's extensive examination produced promising outcomes with the most minor complexity and the best response time.

The subsequent sections of the paper are organized in the following way: section 1 briefly introduces cervical cancer, its causes, and its diagnosis techniques. Section 2 explores the various machine learning techniques and models often used in detecting cervical cancer and is termed a Literature Review. Section 3 comprises a Methodology that explains the flow of the model, Datasets, Confusion Matrix, AUC cove, and Probability Density Curve determining the accuracy of the proposed model. Finally, a conclusion summarizes the paper and our recommendations.

Literature Review

Various ML and DL algorithms were used to detect cancerous cervical cells through pap smear tests and determine our prediction's accuracy. As cervical cancer analysis depends on the knowledge and proficiency of pathologists, who are prone to errors due to their lack of experience in the field, cervical cancer analysis is very subjective.

The study uses KNN, SVM, InspectionV3 along with SVM, Bagging DT, LR, CNN, and Ensemble algorithm, giving accuracies as follows: 79.54%, 90.90%, 99.99%, 96.59%, 79.54%, 90.90% and 87.5%. Among these various models, InspectionV3 and SVM achieved an accuracy of nearly 100%, making it one of the best models among several others.

Many current deep-learning investigations on Pap smear images predominantly concentrate on a binary classification paradigm called two-class classification. Additionally, a prevalent trend involves the analysis of single-cell images rather than the comprehensive examination of raw medical images. The study offers insight into the significance of ML and DL techniques, delineating their evolutionary trajectory within cervical cancer diagnosis, as shown in Fig. 2.

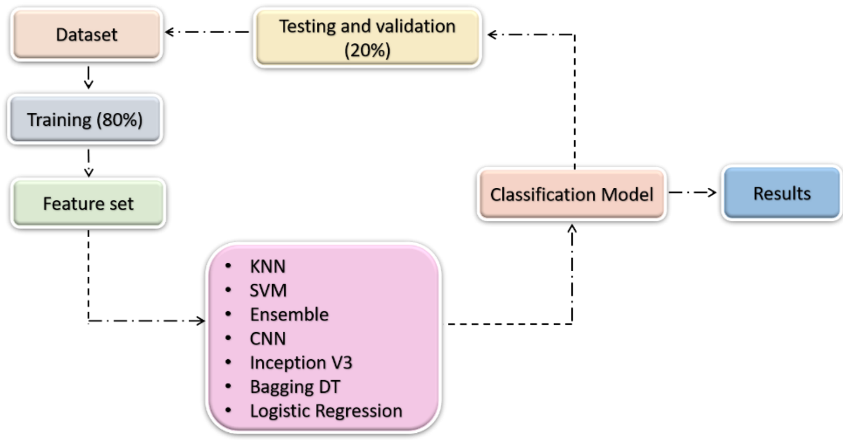


Fig. 2 The generic roadmap for cervical cancer diagnosis.

The application of DL in computer-assisted diagnosis has proven highly beneficial, not only in the realm of cervical cancer but also in broader medical diagnostics. Diagnosing cervical cancer is inherently complex, given its slow progression, underscoring the critical importance of early detection for effective prevention. DL techniques have developed as a valuable tool for pathologists and gynecologists to streamline the diagnostic process. Using interpretable AI methods further contributes by providing meaningful explanations and enhancing pathologists' comprehension. This symbiotic relationship between advanced technology and medical expertise not only aids in accurately identifying cervical cancer but also establishes a foundation for improved diagnostic practices across various medical domains.

Datasets Detail

In this proposed work, three different datasets are used,each providing vital information for the training and testing stages to build and assess ML models for the assessment of cervical cancer. The first dataset used [7] was placed in class 4 in the suggested work analysis. It is among the biggest in this domain and thus offers an extensive compilation of essential characteristics and labels for developing a strong predictive model. Further, the class 2 dataset was identified by the second dataset [8] that was obtained. The predictive model gains accuracy from emphasizing 224x224-resolution cervical cancer screening data, improving its ability to identify early disease signs. Finally, dataset [9], as class 3, was utilized in the work. This Mendeley-sourced dataset adds more dimensions to the study that is being suggested and aids in finding trends and risk factors related to cervical cancer. These datasets' variability ensures a thorough feature exploration, enhancing the precision and flexibility of the proposed ML models for cervical cancer prediction work.

Machine Learning Algorithms

- **SVM**

The main objective of SVM is to differentiate the given data in the best possible way. It is a supervised ML model used in classification and regression. This model is beneficial in various applications, such as in the medical domain for diagnosis and text classification tasks, and it can also handle high-dimension data [10-13].

- **KNN**

KNN can find the resemblance between new and present data, which means that when new data arrives, it can simply classify it into compatible categories. This model is also a supervised learning model used for classification and regression. It has wide applications, such as in the medical field, as it gives clear and understandable methods for prediction models [10].

- **CNN**

Convolutional neural networks are used in image recognition and analysis tasks. Convolutional layers work with filters on input images to gain hierarchical representations. The input layer, hidden layer, output layer, weights, pooling layers, kernels, and filters used in CNN play vital roles at each step to find out even the complex features analysis from a given image. Thus, it is applicable in various applications like detecting cervical cancer, where the tiny details are essential for identification [12-14].

- **Ensemble Algorithm**

In ML, ensemble methods integrate several separate models for a more accurate and reliable prediction framework. Ensemble models are influential as they can decrease overfitting and enhance generalization by combining different viewpoints from distinct models. Ensemble algorithms' predictive performance and tolerance to noisy data are enhanced using strategies such as bagging and boosting. Increased correctness, stability, and adaptation across diverse datasets are prominent benefits, which make them helpful in tackling challenging matters such as cervical cancer estimation [10–11].

- **Logistic Regression**

A popular statistical method in ML for binary classification tasks is LR. Its goal is to predict the chance of an event happening; this probability is often coded as 0 or 1. It is beneficial when the dependent variable indicates a definite result with two classes. The model converts a linear combination of input features into a probability score using the logistic function [10, 15–17] in various disciplines such as finance, medicine, and the social sciences. LR is a valuable and interpretable tool that provides estimations and insights into the likelihood of an event, even though it is clear that LR frequently produces good outcomes, specifically in cases where the attributes and results have primarily linear connections.

- **Bagging Decision Tree**

Using numerous bootstrap samples—random subsets of the original dataset produced by resampling with replacement—many decision trees are trained independently in a Bagging DT model. This method benefits DTs by lowering variance, strengthening the model, and making it more appropriate for larger, more complicated datasets. By integrating these different prediction trees through voting or averaging, Bagging decreases overfitting and enhances the stability and accuracy of the model as a whole. A more thorough and accurate predictive model is produced by bootstrap sampling's varied viewpoints [10, 18–19].

- **Inspection + SVM (Support Vector Machine)**

An image block model is the objective of the Inception Model, a DL model, to give the optimal local sparse structure possible in an SVM. Instead of being restricted to using only one filter size, which can be concatenated before moving on to the next layer, it enables us to use several filter sizes in a single image block.

Proposed Methodology

This paper's main objective is to investigate and understand the process, including timely and precise identification of invasive cells in the cervix. Cell classification allows for the identification of cell anomalies, the detection of the precancerous stage, and the possibility of an early analysis to control the progression of the disease. This paper uses an approach known as transfer learning—the capacity to adapt skills and information from previous work to new tasks—to classify the cervical classes. As seen in Fig. 3, it is a standard method to automatically extract features using pre-trained models from a new dataset. Further, it is noticed that the prediction from the Inception and SVM gave us an accuracy of nearly 99.99%, which could be seen from the probability density graph. The usage of "inception modules," which are modules made up of numerous convolutional layers with various kernel sizes and pooling procedures, is what distinguishes the Inception architecture. These modules are designed to efficiently capture features at different spatial scales, allowing one to learn diverse features from the input images. Its primary motivation is to increase the suggestive power of SVM while keeping computational costs manageable.

Assume a binary classification task with two classes (labeled +1 and -1). This work uses a training dataset with matching class labels (Y) and input feature vectors (X). Equation (1) for the linear hyperplane can be written as follows:

$$w^T+b=0, \quad (1)$$

Vector w signifies the direction perpendicular to the hyperplane, termed the average vector. The b parameter in the equation denotes the distance, or offset, of the hyperplane from the beginning along the normal vector w.

As demonstrated in equation (2), the distance between the decision boundary and data point x_i can be computed as

$$d_i=(w^t*xi + b)/||w||, \quad (2)$$

where the weight vector w's Euclidean norm is denoted by ||w||, the norm of the standard vector is in Euclidean space. Equation (3) for Linear SVM classifier

$$y=\{ \begin{array}{l} 1 : w^t x+b \geq 0 \\ 0 : w^t x+b<0 . \end{array} \quad (3)$$

With the combination of ML and DL techniques, the proposed system has developed an integrated model with the following features:

- **Feature Extraction with Inception:** The Inception framework is used as a feature extractor, and instead of training the entire Inception network from scratch, the pre-trained Inception framework is often used. The pre-trained Inception model is run on the input images, and the output of one of the intermediate layers is extracted. These outputs serve as high-level features that capture the foremost traits of the input images.

- **Data Filtering** refers to selecting or excluding specific data points from a dataset based on predefined criteria. It is a common data preprocessing technique used to clean, refine, or reduce the size of a dataset before performing further analysis or modeling. It has several features, such as noise removal, data reduction, quality control, and feature selection, employed using machine learning techniques.
- **SVM for Classification**: An SVM classifier uses the retrieved features as input. For classification problems, supervised learning algorithms like SVMs are employed. They work by identifying the feature space hyperplane that most effectively divides the different classes. SVMs work very well with high-dimensional data, which makes them ideal for jobs where the number of features might be huge, such as image classification.
- **Training and Fine-Tuning**: Reliant on the specific work and dataset, the SVM classifier may be trained directly on the extracted features or fine-tuned along with the extracted features using a dataset. Fine-tuning involves adjusting the parameters of the SVM to suit better the specific task or dataset, which can lead to improved performance.
- Once the SVM classifier is trained or fine-tuned, it can classify new images. The performance of the combined Inception-SVM model is examined on a separate test dataset.

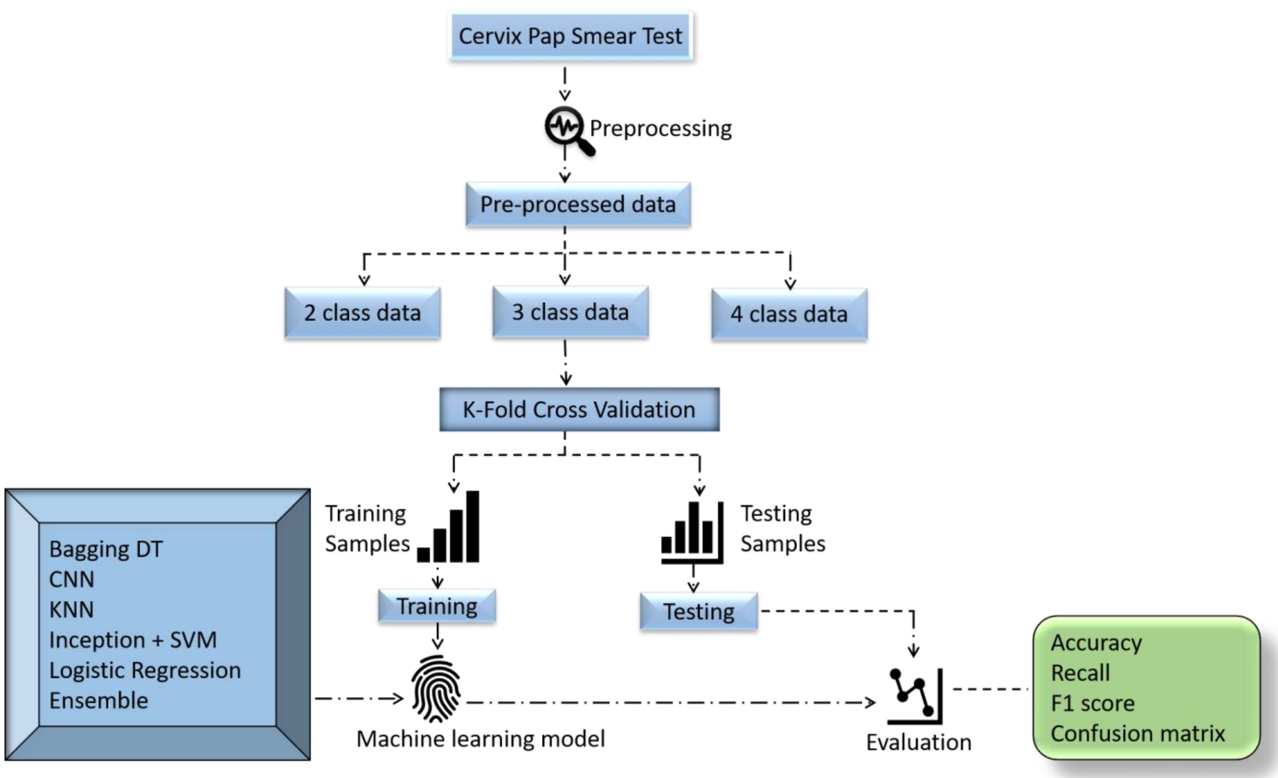


Fig. 3 Schematic illustration of the suggested workflow for cervical cancer diagnosis.

A supervised ML approach for regression and classification problems is called SVM. It works very well in high-dimensional areas for classification jobs—the mathematical derived formula for SVM. For the sake of simplicity, a pseudo code for instructing the working of various ML models is given.

To assess its accuracy, precision, recall, and other relevant metrics. This combination has enhanced the image classification task by removing specific features from the images provided in the dataset. The concept of epochs and the associated training and validation loss values over epochs are specific to iterative models, typically neural networks. Traditional ML models such as SVM, DTs, LR, etc., do not use the concept of epochs. Fig. 4 shows the model accuracy graph to epochs.

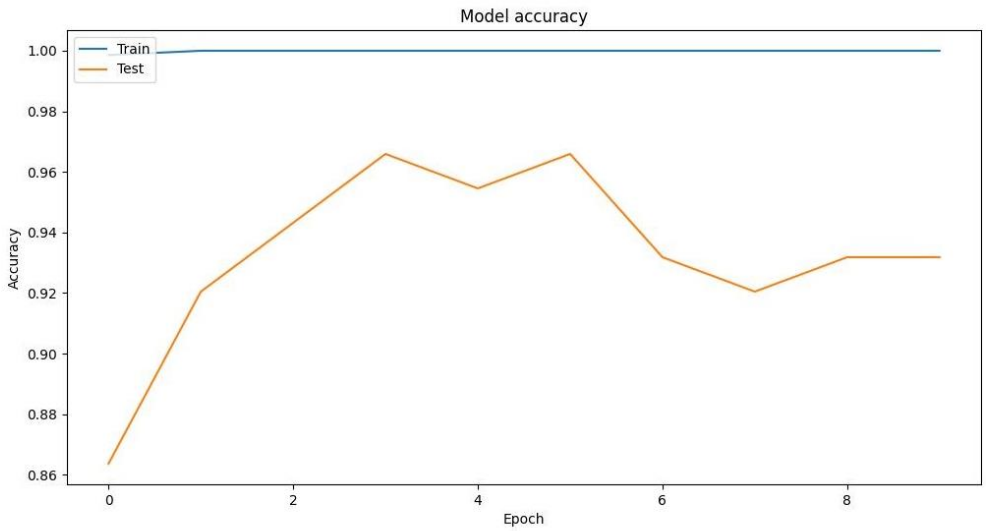


Fig. 4 Model accuracy graph to epochs.

They are trained in one shot using the entire dataset and do not provide a history of loss values over epochs.

Procedure

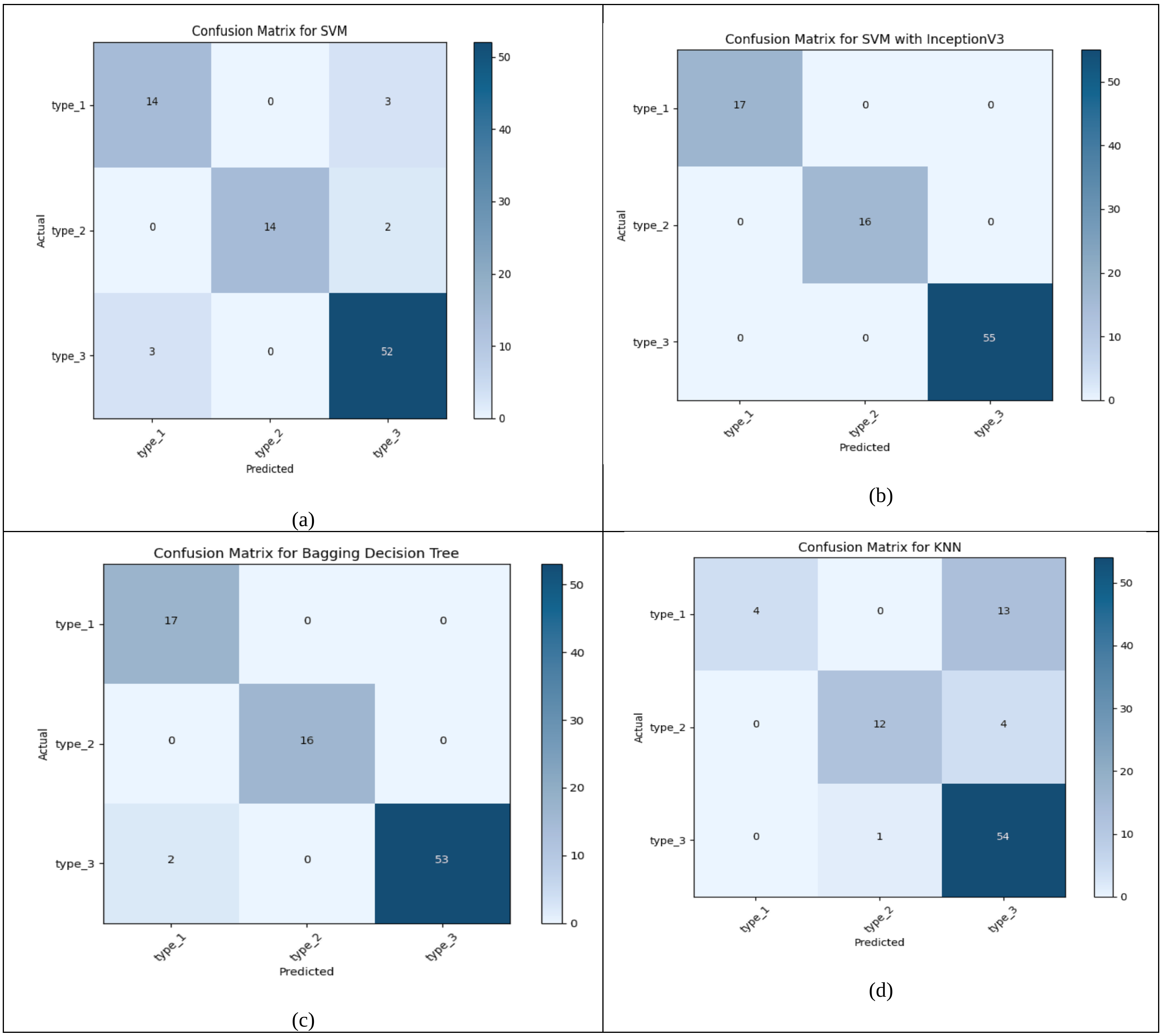
1. Preprocessing
    - Adjust the image of 256 by 256.
    - Set every input's scale to a value between 0 and 1.
    - Split training test with sklearn.model_selection.
2. Build Models
    - Create all the different models in this step. Namely SVM, LR, Bagging DT, Inception +SVM, etc
    - Initialize the Model () of each ML technique that one wants to use
    - Now, train the models using the training data
    - Make predictions using the test data, then estimate the accuracy of these given models' forecasts by comparing them with the actual labels.
3. Model Fitting
    - (X_train, y_train)
    - Validation_data = (X_val, y_val),
    - Model callbacks=[tensorboard_callback])
4. Assessment of the Model
    - model.predict(X_test)
5. Confusion matrix
    - confusion_matrix(y_test, y_predict.argmax(axis = 1))
6. Determine the metrics for evaluation
    - For each y_predict, do
    - Determine the specificity, accuracy, sensitivity, F1-score, and recall.

    End.

Experimental Analysis and Findings

Because cervical cancer cell classification is complex, using several ML approaches has become essential as it provides flexibility and improved accuracy. The DL approach splits the problem into several pieces, finds solutions for each, and delivers findings considering every response in detail. This probably helped the DL model perform better in the learning system's cervical cancer classification test. Simultaneously, ML divides the problem into several components, finds solutions for each, and simply sums the outcomes. Fig. 5 shows the confusion matrix for the machine learning techniques used in the proposed work.

The test approach is trained and validated using the suggested model. This is done by applying the ratio of 80:10:10. In this ratio, 80% of the data received is used to teach the proposed ML algorithm to make the model more accurate. In comparison, 10% of the data is used for validation purposes to train the artificial intelligence model for finding an optimized approach to solve the given problem, and the remaining 10% is used for testing purposes to ensure that the models function as intended and produce dependable results when deployed in a real-time world. The precision of the model is presented in Fig. 6.



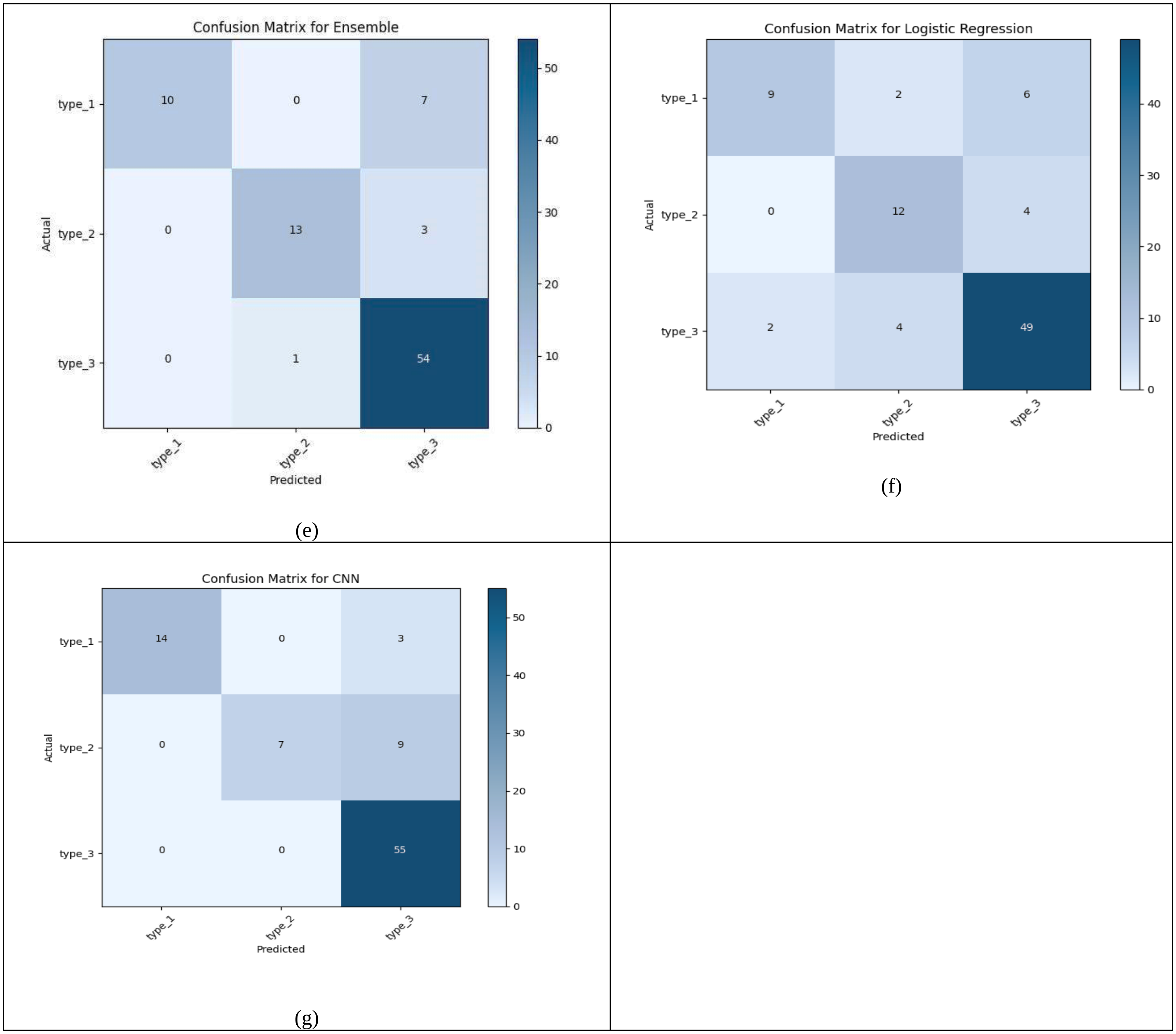


Fig. 5 Illustrate the confusion Matrix for (a) SVM, (b) SVM+InceptionV3, (c) Bagging Decision Tree, (d) K- Nearest Neighbor, (e) Ensemble, (f) Logistic regression, (g) CNN.

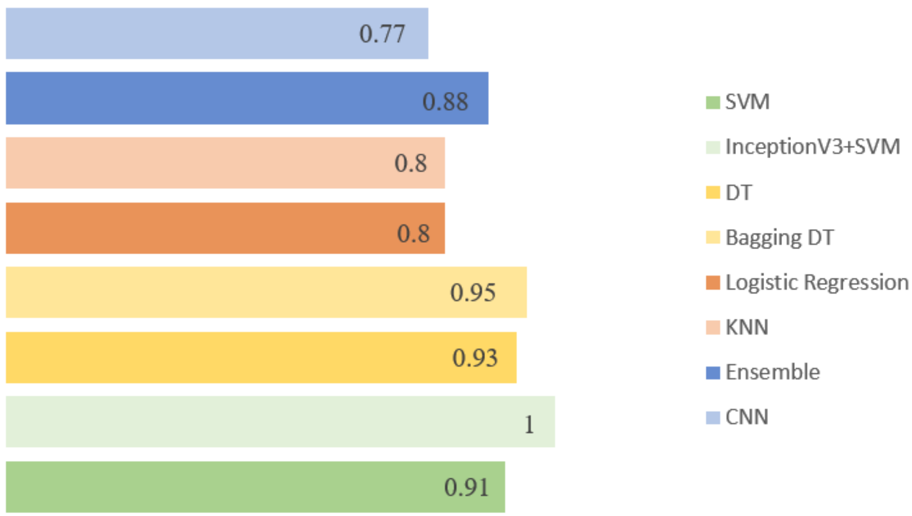


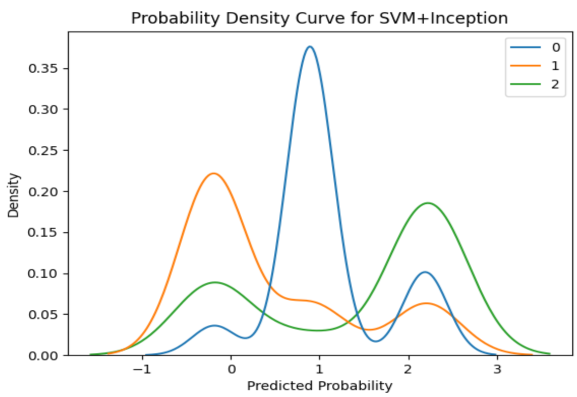
Fig. 6 Accuracy of various models.

The model is trained and validated using about 3000 training images divided into four classes.
The depth of the layer, Probability Density Curve, initial learning rate, Confusion matrix, and AUC cove are the numerous parameters that determine the model statistically.
This study assessed and compared the results of the SVM and Inception models with those from all other ML models employing transfer learning.

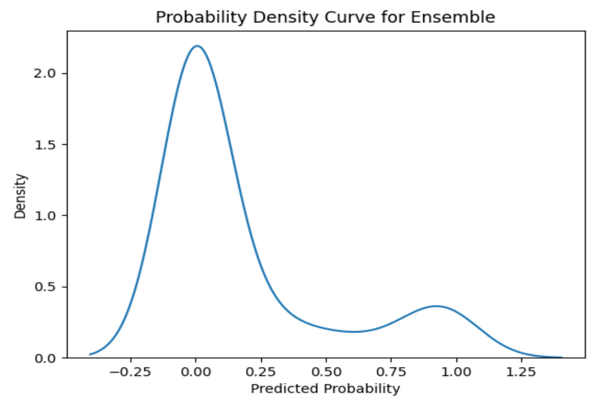
The performance comparison graph is provided for the proposed models. Color shading allowed the best results to be easily identified as the highest-performing Inception + SVM. Apart from the graphical representation for each model, the suggested work compared models using several parameters, such as precision, F1 scores, and recall. The ratio of what is true to what is deemed accurate is called precision, or PPV. Recall, sometimes referred to as sensitivity, is the ratio of what is true to what the model predicts to be true. The harmonic mean of recall and precision is termed the F1-score.

Precision = TP / (TP + FP) (5)
Recall = TP / (TP + FN) (6)
F1 Score = (2 ∗ precision ∗ recall) / (precision + recall) (7)
Accuracy = (TP + TN)/ (TP + FN + TN + FP) (8)

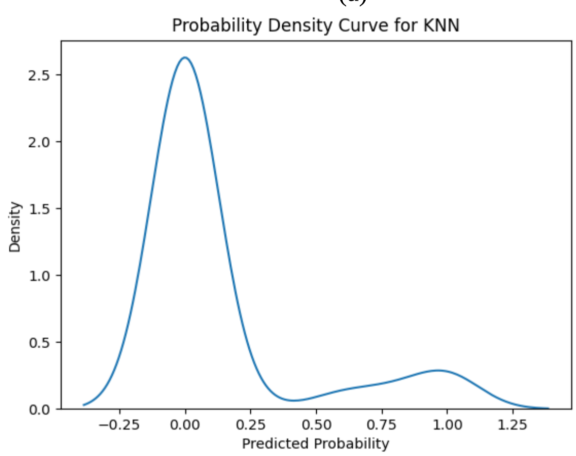
The metrics for each model are integrated with the cell classifications to establish the general robustness of the pre-trained Inception+SVM model for multi-class cervical cancer classification. This will facilitate our comprehension of the relationship between classification performance, class distribution, and cell type. A confusion matrix has been presented to assess the precision and utility of the framework of different classes. Also, their graphical representation for the predicted probability density curve is illustrated in Fig. 7.


Probability Density Curve for SVM+Inception
Density
Predicted Probability
0
1
2

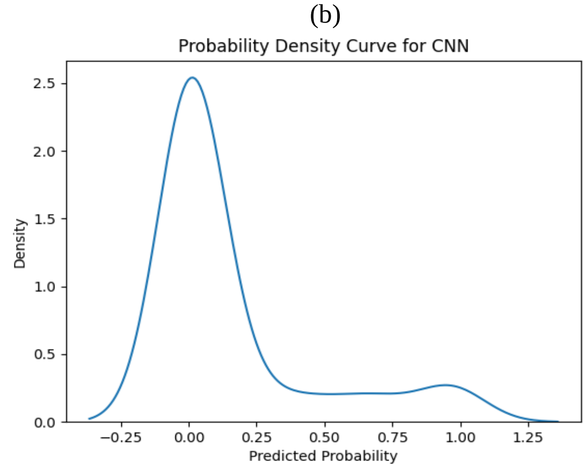
(a)


Probability Density Curve for Ensemble
Density
Predicted Probability

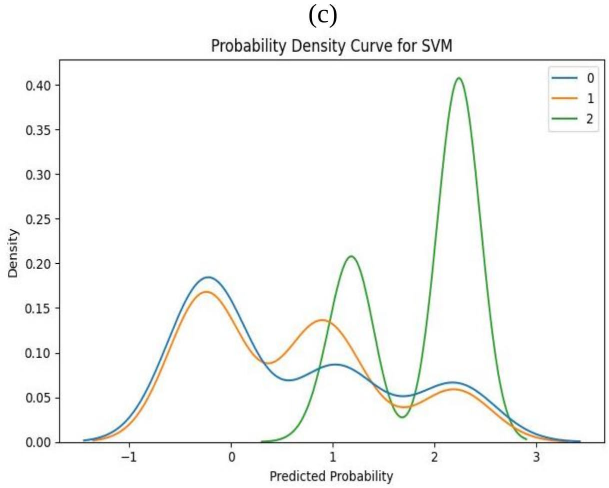
(b)


Probability Density Curve for KNN
Density
Predicted Probability

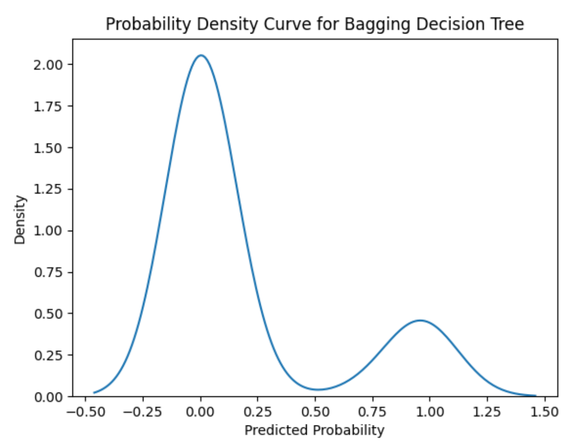
(c)


Probability Density Curve for CNN
Density
Predicted Probability

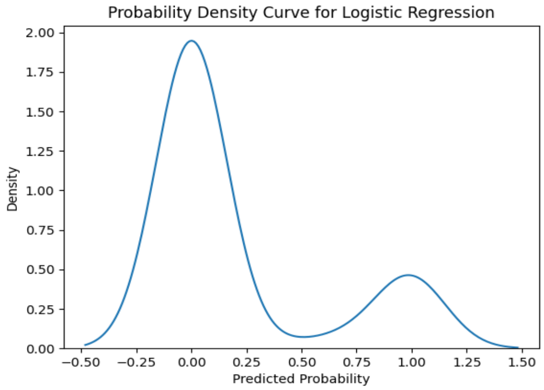
(d)


Probability Density Curve for SVM
Density
Predicted Probability
0
1
2

(e)


Probability Density Curve for Bagging Decision Tree
Density
Predicted Probability

(f)



(g)

Fig. 7 shows the predicted probability density curve for the various models (a) SVM+Inception, (b) Ensemble, (c) KNN, (d) CNN, (e) SVM, (f) Bagging Decision Tree, and (g) Logistic Regression.

The informative elements taken out of the dataset are the leading cause of the high precision of the suggested technique. Due to the collected deep features, the machine learning classifier is better able to predict and validate cervical cells. Table 1 shows the parameters used in the proposed work for various models. Fig. 8 shows the graphical representation of multiple models for (a) accuracy, (b) Recall, and (c) F1-Score.

Table 1 depicts various parameters for several models.

Model	Precision	Recall	F1-score
SVM	0.911937	0.881328	0.895145
InceptionV3+SVM	1	1	1
DT	0.929187	0.920544	0.922559
Bagging DT	0.945304	0.981818	0.962035
Logistic Regression	0.771786	0.72344	0.73613
KNN	0.894547	0.655704	0.68856
Ensemble	0.892977	0.788124	0.826419
CNN	0.944444	0.77451	0.826328

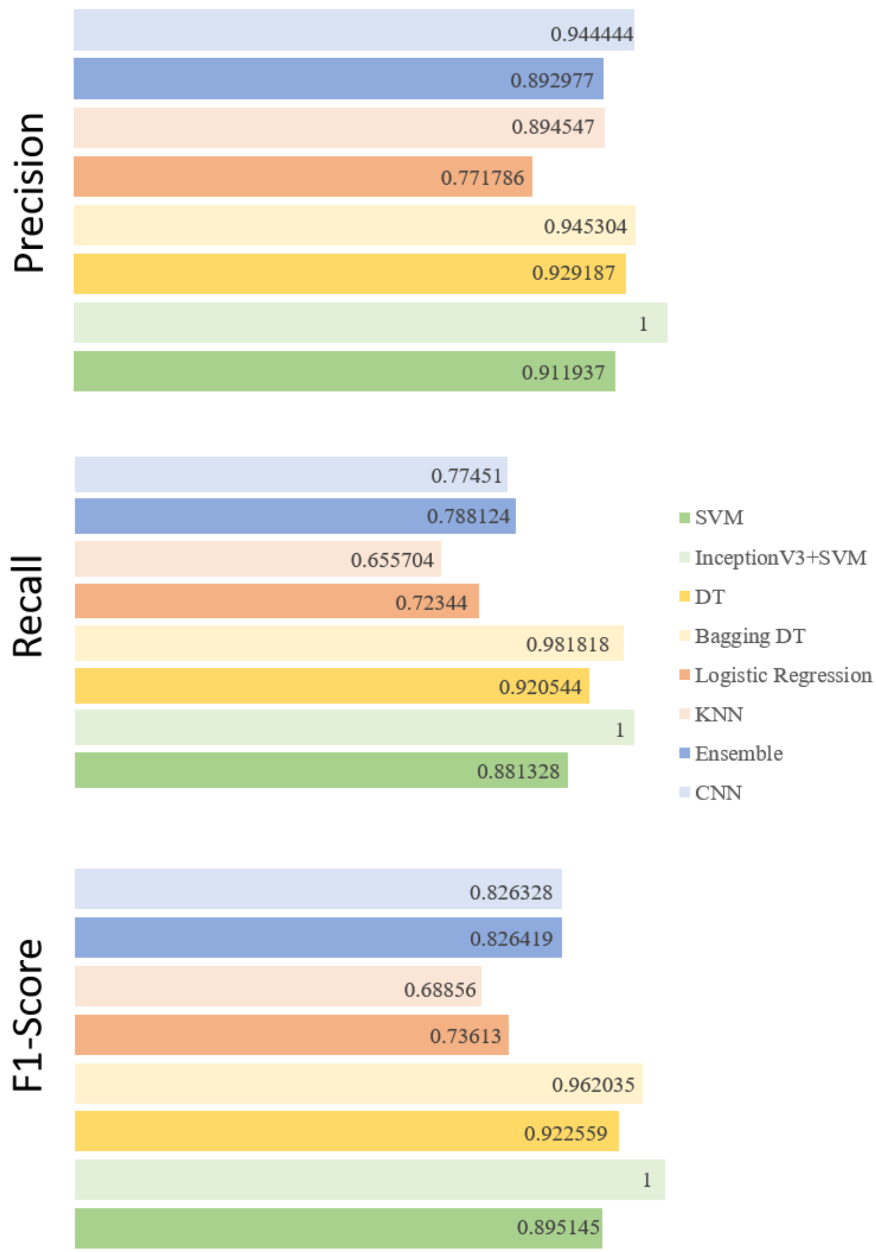


Fig. 8 Graphical representation of various models for precision, Recall, and F1-Score.

Conclusion

Because of their intricate architecture, cervical cells must be examined in person over several hours in a laboratory. Due to the rise in cervical cancer incidence, it is now essential to diagnose the disease in its precancerous or early stages and to cut down on the time and resources used for the diagnosis process to minimize associated overhead costs. The suggested model is trained by the data used in this work, a publically accessible dataset of cervical cell pictures, to categorize cell images into four main cell types.

Health professionals can predict unusual cervical cells that could lead to cancer by using this classification to separate normal cells from abnormal cells. A 99.9% overall testing accuracy was obtained with this effective model. Aspects of DL and ML are combined in the proposed Inception + SVM model for cervical cancer classification. The most complex and nuanced features from images may be retrieved with DL. Researchers can improve this model by utilizing various DL approaches to produce computationally promising findings more quickly. For the model to provide a better degree of generality in the future, it has to be trained on a more extensive set of Pap smear images.

New deep architectures, such as ResNet and tree-based techniques, have recently demonstrated positive outcomes in several applications. These architectures may be investigated in the future to find cervical cancer.

Declaration

Conflict of interest: The authors have no conflict of interest.
Author contributions Pandey A, Arora J, Kohli GSR, and Khurana M contribute to experimental analysis and findings. Sharma N and Jindal N reviewed the manuscript.
Availability of data and materials Not applicable.
Funding None.

References

1. Kalbhor M, Shinde S, Joshi H, & Wajire P. Pap smear-based cervical cancer detection using hybrid deep learning and performance evaluation. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 2023, 11(5):1615-1624. DOI:10.1080/21681163.2022.2163704
2. Alsubai S, Alqahtani A, Sha M, Almadhor A, Abbas S, Mughal H, & Gregus M. Privacy Preserved Cervical Cancer Detection Using Convolutional Neural Networks Applied to Pap Smear Images. Computational and Mathematical Methods in Medicine, 2023. DOI: 10.1155/2023/9676206
3. Rivers G, Hinchliff S, & Thompson, J. Transgender and non-binary people's experiences of cervical cancer screening: A scoping review. 2024, Journal of Clinical Nursing.
4. Chandran V, Sumithra M G, Karthick A, George T, Deivakani M, Elakkiya B, … & Manoharan S. Diagnosis of cervical cancer based on ensemble deep learning network using colposcopy images. BioMed Research International, 2021.
5. Fong AWM, Assumpção PPD, Valadares LJ, & Moreira FC. Microbiota changes: the unseen players in cervical cancer progression. Frontiers in Microbiology,2024, 15, 1352778.
6. Ghoneim A, Muhammad G, & Hossain MS. Cervical cancer classification using convolutional neural networks and extreme learning machines. Future Generation Computer Systems, 2020, 102:643-649.
7. https://www.kaggle.com/datasets/prahladmehandiratta/cervical-cancer-largest-dataset-sipakmed.
8. https://www.kaggle.com/datasets/ofriharel/224-224-cervical-cancer-screening
9. https://data.mendeley.com/datasets/zddtpgzv63/4
10. Tandel GS, Balestrieri A, Jujaray T, Khanna NN, Saba L, & Suri JS. Multi-class magnetic resonance imaging brain tumor classification using artificial intelligence paradigm. Computers in Biology and Medicine, 2020, 122, 103804.
11. Ang DJM, & Chan JJ. Evolving standards and future directions for systemic therapies in cervical cancer. Journal of Gynecologic Oncology, 2024, 35(2).
12. Wang M, Huang K, Wong MC, Huang J, Jin Y, & Zheng Z J. Global cervical cancer incidence by histological subtype and implications for screening methods. Journal of Epidemiology and Global Health, 2024, p.1-8.
13. Tan SL, Selvachandran G, Ding W, Paramesran R, & Kotecha K. Cervical Cancer Classification from Pap Smear Images Using Deep Convolutional Neural Network Models. Interdisciplinary Sciences: Computational Life Sciences, 2023, p.1-23. DOI: 10.1007/s12539-023-00589-5
14. Ma K, Chen Y, Zhang X, Zheng H, Yu S, Cai S, & Xie L. sEMG-based trunk compensation detection in rehabilitation training. Frontiers in Neuroscience, 2019, 13, 475642.
15. Kalbhor M, Shinde S, Joshi H, & Wajire P. Pap smear-based cervical cancer detection using hybrid deep learning and performance evaluation. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 2023, 11(5):1615-1624.
16. Basak H, Kundu R, Chakraborty S, & Das N. Cervical cytology classification using PCA and GWO enhanced deep features selection. SN Computer Science, 2021, 2(5): 369.
17. Jia AD, Li BZ, & Zhang CC. Detection of cervical cancer cells based on strong feature CNN-SVM network. Neurocomputing, 2020, 411:112-127. DOI: 10.1016/j.neucom.2020.06.006
18. Sharma R, & Kukreja V. Image segmentation, classification and recognition methods for comics: A decade systematic literature review. 2024, Engineering Applications of Artificial Intelligence.
19. Sharma R, & Mehta K. Credit risk and bank specific factors: an empirical study using panel GMM. International Journal of Economics and Business Research, 2024, 27(2):294-309.