# A Hybrid Ensemble Learning Approach for Cervical Cancer Detection: Combining Multiple CNN Models for Enhanced Diagnostic Accuracy

Aishwarya N. Kumar \* and Meenakshi Sundaram A.

School of Computer Science and Engineering, REVA University, Bengaluru, India Email: aishwaryankumar2606@gmail.com (A.N.K.); meenakshi.sa@reva.edu.in (M.S.A.) \*Corresponding author

Abstract—Cervical cancer is now the third most prevalent type of cancer in the universe. An early-stage detection may result into surgical procedure called hysterectomy. The majority of these cases are associated with the risk of infection from Human Papilloma Virus (HPV). Preventive measures, while the costliest approach to cancer prevention, can safeguard approximately 37% of cases. The Pap smear is a routine diagnostic tool used for the initial screening of cervical cancer. However, this manual procedure often results in a high number of false positives due to human error. Currently, data mining-based concept has gained huge attention in this domain of predictive analysis for disease detection where Machine Learning (ML) based models are widely adopted to predict the cervical cancer where supervised ML methods have played significant role. However, the performance of these models is affected for large dataset and computational complexity related issues also become more challenging. To address the issues of ML, researchers have introduced deep learning-based method to enhance the pattern learning capability of classification models to improve the overall accuracy. however, the blackbox nature of these models can affect the system performance therefore combing two or multiple models can be beneficial to avoid this risk. In this work, we present a hybrid deep learning model where three Convolutional Neural Network (CNN) models are combined together and an averaging model is used to make the final decision related to prediction. Evaluated on a separate test set, the ensemble approach demonstrates improved diagnostic performance, achieving a higher accuracy compared to individual models. This study underscores the effectiveness of combining multiple CNNs in ensemble learning to advance the accuracy and reliability of cervical cancer detection, offering a promising tool for early diagnosis and improved patient outcomes. The experimental analysis shows that the average classification accuracy is obtained as 0.95, 0.91, 0.95, and 0.96 by using Random Forest (RF), Support Vector Machine (SVM), Decision Tree (DT), and proposed model, respectively.

*Keywords*—cervical cancer, Human Papilloma Virus (HPV), Convolutional Neural Network (CNN), ensemble approach, Random Forest (RF), Support Vector Machine (SVM), Decision Tree (DT)

## I. INTRODUCTION

A Cancer is identified as one of the complex diseases marked by the unrestrained growth and multiplication of abnormal cells, causing from mutations in DNA [1]. These cells can invade healthy tissues, forming tumors that may cause significant damage and, in some cases, lead to fatal outcomes [2]. Generally, these tumors are categorized as malignant and benign tumors [3]. Malignant tumors can spread to surrounding tissues or distant parts of the body, creating new tumors through a process known as metastasis [4]. In contrast, benign tumors are not hazardous for the neighboring tissues [5]. One of the greatest challenges for clinicians is distinguishing between these tumor types before the disease becomes advanced. The severity of cancer and the effectiveness of treatment largely depend on the specific type of cancer and the timing of its diagnosis.

In 2018, cancer was identified the most common cause of 9.6 million deaths worldwide. Therefore, cancer is reported to be the second most common cause of death worldwide. According to the International Agency for Research on Cancer (IARC), 1 in 6 deaths globally is attributed to cancer [6]. A report from the World Health Organization (WHO) underscores that cancers predominantly affecting women, such as breast, cervical, and ovarian cancers, are leading contributors to premature mortality among women globally [7]. Cervical cancer alone accounts for the deaths of over 270,000 women annually, with more than 85% of these fatalities occurring in developing nations. According to the IARC data 2018, the global impact is evident in the estimated 444,500 new cervical cancer cases reported each year [8].

While the burden of cervical cancer has significant impact on developing countries, high-income nations like the United States and England are also grappling with a concerning rise in cases. Recently, American Cancer Society presented a report in 2019 which shows that in the U.S., around 13,420 new cervical cancer diagnoses are expected annually, alongside approximately 4,170 deaths [9]. Similarly, in England, the incidence rate of cervical cancer among women has escalated from 2.7% to 4.6%. On a global scale, it is projected that over 527,624

Manuscript received October 17, 2024; revised January 2, 2025; accepted March 3, 2025; published June 19, 2025.

new cases of cervical cancer and 1,671,149 new cases of breast cancer are identified each year [10]. In India, the situation is particularly dire where total 122,844 new cervical cancer cases and 144,937 new breast cancer cases were reported annually. India has accounted  $1/3^{rd}$  of the total cervical cancer deaths worldwide. Indian women face a 1.6% lifetime risk of developing cervical cancer, with a 1.0% risk of dying from the disease [11]. Breast cancer statistics are similarly troublesome, with a cumulative incidence of 2.7% and a mortality rate of 1.5% [12].

According to GLOBOCAN 2020 data, cervical cancer ranks as the third most common cancer in India, with an incidence rate of 18.3% with total 123,907 cases and it has reported the mortality rate of 9.1% [13]. The agestandardized incidence rate per 100,000 population stands at 18, while the five-year prevalence rate across all age groups is 42.82 per 100,000. The National Cancer Registry Programme identifies breast and cervical cancers as the most prevalent cancers among Indian women, with cervical cancer representing 6-29% of all female cancer cases in the country. Notably, the district of PapumPare in Arunachal Pradesh has the highest incidence rate of cervical cancer in Asia, with 27.7 cases per 100,000 populations [14]. In this work, we mainly focus on the cervical cancer. The early detection of this can lead to reduce the mortality rate. Cervical cancer develops in the cervix, the lower part of the uterus, and is one of the leading causes of cancer-related deaths among women. Although all women are at risk, the disease most frequently affects those aged 30 and older. The primary cause of cervical cancer is the Human Papilloma Virus (HPV), a virus transmitted through sexual contact. HPV is highly prevalent, infecting at least half of sexually active individuals at some point in their lives. One of the primary surgical interventions for cervical cancer, particularly in its early stages, is a hysterectomy. A hysterectomy involves the removal of the uterus and, in certain cases, surrounding tissues and organs such as the cervix, ovaries, and fallopian tubes [15]. However, cervical cancer is largely preventable [16]. A highly effective vaccine exists that protects against the most dangerous strains of HPV, significantly reducing the risk of developing the disease. Additionally, early detection through regular screening tests, such as Pap smears and HPV testing [17], can help identify precancerous changes in the cervix. When detected early, cervical cancer responds exceptionally well to treatment, resulting in high recovery rates.

Currently, we have noticed a tremendous growth in technological advancements which has led to improve the efficiency in medical domain. For instance, a huge amount of data is collected regularly and made accessible to the medical research community to improve the accuracy in disease prediction and diagnosis. Currently, research community is continuously working on improving the efficiency of data analysis models which are used to uncover the valuable insights and aiming to use this mined information to support diagnostic decision-making processes. Machine learning has been increasingly applied to tasks such as predicting and detecting cervical cancer at an early stage. Advances in technology, when integrated with machine learning techniques, present an opportunity to address cervical cancer through a more comprehensive approach. Machine learning experts are dedicated to refining predictive models that can process large datasets related to cervical cancer, identifying and extracting crucial information to aid in diagnostic decision-making. Several machine learning-based models have been introduced to detect the cervical cancer such as Support Vector Machine (SVM) [18], combined machine and deep learning [19]. Some of the recent methods are discussed in Section II.

However, the performance of these Machine Learning (ML) and Deep Learning (DL) based systems is affected due to several parameters such as high computational cost, lack of dimensionality management, inaccurate handling of imbalanced data, poor accuracy, and precision. Therefore, performance of these models needs to enhanced further. The previous discussion has reported that several researchers have focused on DL models because of their capacity to learn complex data patterns. Therefore, in this work also, we focus on development of novel deep learning approach and introduced a deep ensemble learning model where different DL models are combine together to obtain the robust classification decision. The main contribution of this work are as follows:

- First of all, we focus on data pre-processing phase, removing empty data, finding missing values, imputing missing values and normalization are performed.
- The pre-processed data is then fed to the proposed hybrid CNN classifier where three CNN models are used to learn the complex patterns of the data.
- Later, an averaging method is present to obtain the final prediction.

In order to achieve these objectives, the proposed approach focus on following research questions:

- RQ 1: How can advanced deep learning models improve the accuracy and robustness of cervical cancer detection?
- RQ 2: What pre-processing techniques are most effective for handling missing and noisy data in cervical cancer datasets?
- RQ 3: Can an ensemble of CNN-based models provide superior classification performance compared to standalone models?

The remainder of this article is structured as follows: Section II provides a brief literature review, discussing existing methods. Section III introduces the proposed deep ensemble learning model for cervical cancer detection. Section IV presents the results of the proposed model and compares its performance with current approaches. Finally, Section V offers concluding remarks and outlines the future scope of this research.

#### II. LITERATURE REVIEW

In this section, we discuss about the existing methods of cervical cancer detection based on the concept of machine and deep learning methods. Recently, Gnanavel *et al.* [20] reported that the existing methods, i.e., Pap smears are

time consuming and prone to errors therefore suggested the importance of automation. Therefore, authors addressed these issues by introducing explainable Artificial Intelligence (XAI) techniques, including GradCAM, GradCAM++, and Layer-wise Relevance Propagation (LRP), to enhance the transparency and reliability of cervical cell classification models. Using the Herlev Dataset, the research applies data pre-processing and augmentation strategies to develop a binary classification model, achieving a 91.94% accuracy with VGG16. XAI-based qualitative analysis revealed that the model primarily focused on nucleus and cytoplasm features, which are critical indicators of malignancy. Quantitative evaluation showed LRP to be the most effective technique, with the lowest mean image entropy (2.4849) and significant prediction confidence drops under perturbations, solidifying its role in improving cervical cancer detection interpretability.

Kumawat et al. [21] focused on supervised ML based approach for this task of cervical cancer classification. In this work, authors have performed two experiments with and without feature selection mechanism and later, these obtained features are classified with the help of six different classification methods. The feature selection is performed by using relief rank, wrapper and Least Absolute Shrinkage and Selection Operator (LASSO) regression methods and the obtained features are classified with the help of artificial neural network, Bayesian network, SVM, random tree, logistic tree and Expected Goals (XG) Boost tree. The XG boost has reported the high accuracy of 94.094% without integrating any feature selection mechanism. However, these methods reported several limitations such as overfitting, poor interpretability, and accuracy. Ali et al. [22] presented ML based approach for early detection of cervical cancer. The dataset used in this work consist of biopsy, cytology, Hinselmann, and Schiller. The obtained feature set is then transformed with the help of log, sine function, and Z-score methods. Finally, supervised classification methods were applied where Random Tree classifier is used which has reported the accuracy of 98.33% and 98.65% for biopsy and cytology data.

Habtemariam et al. [23] presented computer visionbased model for Cervix Type and Cervical Cancer Classification by using deep learning method. This model uses histopathology image data and a lightweight MobileNetv2-YOLOv3 pre-trained model is used to extract the Region of Interest (ROI). Then the obtained ROI images are fed to the final classification model where pre-trained EfficientNetB0 was trained to classify the histopathological images. Tripathi et al. [24] presented deep learning based framework where ResNet-152 architecture is used and it has reported the average accuracy of 94.89%. Chen et al. [25] developed CytoBrain approach for cervical cancer detection. This approach is carried out in three main phases where first of all cervical cell segmentation is performed, later cell classification task is carried out and finally human aided diagnosis is incorporated. This model uses whole slide images and uses compact Visual Geometry Group (VGG)to perform the cell classification. Chandran et al. [26] focused on colposcopy for cervical screening and discussed its importance in lowering the incidence and mortality rate from cervical cancer. However, the visual screening leads to lower diagnostic efficiency and misdiagnosis. Therefore, authors of this work introduced DL based computer visionbased approach by considering colposcopy images. This work considers two CNN based architectures as VGG19 as transfer learning and CYENET which is used to develop the Colposcopy image classification. The VGG19 model has reported the overall accuracy of 73.3% whereas the ensemble CYENET model has reported the average accuracy of 92.3%. Kalbhor et al. [27] reported that the existing Pap-smear test suffer from the high rate of falsepositives therefore authors have suggested to develop computer aided diagnostic system based of deep learning framework. Therefore, this article presents fuzzy min-max neural network-based neuro-fuzzy architecture and combined deep learning model to generate a hybrid DL architecture. The DL model uses AlexNet, ResNet-18, ResNet-50 and GoogleNet model. The combination of ResNet-50 with AlexNet has reported the highest accuracy of 95.33%. Alyafeai and Ghouti [28] used pre-trained deep learning model for cervical tumor image classification. the main aim of this work is to perform cervix detection and cervical classification. The first phase reported the detection accuracy as 0.68 in terms of IoU. In the next phase, the self-extracted features are learned using lightweight CNN where it has reported the AUC of 0.82.

Jahan et al. [29] presented an analysis of various ML algorithms such as random forest, gradient boosting, logistic regression etc. for cervical cancer classification based on the concept of data mining. However, these datasets suffer from missing values which has impact on classification performance. therefore, missing value imputation methods are used. Further, this work focuses on selection of best features therefore Chi -square, SelectKBest methods were used. Finally, Random Forest classifier is used to train and classify the data. In this work, the MLP classifier has reported the highest accuracy. Lu et al. [30] presented an ensemble approach to predict the cervical cancer with the help of voting scheme. The complete model uses data correction, missing value imputation, feature extraction and classification. The classifier model uses SVM, K-Nearest Neighbors (KNN), Multilaver Perceptron (MLP) and Decision Tree model to train the ML model for classification.

## III. MATERIALS AND METHODS

This section presents the detailed discussion about proposed model for cervical cancer classification. as discussed before, a significant amount of work has been carried out so far where machine learning based systems have been adopted widely because it has several advantages over traditional methods such as complete automated approach, no human intervention thus no human error, scalable to huge data size etc. however, the machine learning based methods also face several challenges when implemented for these tasks. Therefore, to overcome the issues of ML based approaches, researchers have developed deep learning based models and suggested to incorporate these DL based model to achieve improved classification accuracy. therefore, this work focuses of development of deep learning-based model for cervical cancer classification.

## A. Overview of Proposed Approach

The proposed approach of cervical cancer classification is based on the concept of data mining where supervised learning is applied to predict the categorical labels for each instance in the given dataset. It is a process to map the input features to the discrete output class. Classification is a supervised learning task in data mining where the goal is to predict a categorical label for each instance in a dataset. It involves building models that map input features to discrete output classes. The complete data mining-based classification approach follows a structured process such as data collection, data analysis, data preparation, modelling, and performance evaluation. The data collection phase includes collecting the data from the corresponding domain, perform exploratory analysis, accessing data quality by analysing the missing values, outliers, noise, and inconsistencies in the data. The data preparation phase prepares the data for modelling phase by

performing cleaning, transforming and organizing it. This phase includes data cleaning by imputing the missing values, outliers, removing duplicates etc. in next step, feature engineering and feature selection tasks are performed where data encoding, applying domain knowledge and selecting the relevant features tasks are performed. Finally, the data is split into training data, testing data and validation data. The next phase includes the building the model where classifier model is built and trained based on the selected attributes resulting in producing a trained model. Finally, this trained model is used to test the performance of the system and performance of model is evaluated based on several performance measurement parameters. Below given figure depicts the complete architecture of proposed model. In this work, we have considered the publicly available dataset for cervical cancer detection. The dataset consists of missing values therefore, we applied missing value imputation method, later, feature selection process is applied to select the best attributes. The obtained features are then fed to the CNN based ensemble classifier where majority voting mechanism is applied to obtain the final outcome of classifier model.



Fig. 1. Complete architecture of proposed model.

#### B. Missing Value Imputation

The missing values in the dataset affects the classification performance thus a missing value imputation method can be beneficial to handle these values and provide suitable values to enhance the overall classification performance. Generally, these data samples are obtained from various resources thus it becomes vulnerable to different types of noise, missing values and outliers. To overcome these issues, data pre-processing becomes important factor in data mining-based classifier models and it plays important role in knowledge discovery. *n* this dataset, some attributes have incomplete data for instance "Time since first diagnosis" and "Time since last diagnosis" have 92% of missing values therefore we remove these missing values. Similarly, the cervical and condylomatosis Acquired Immunodeficiency Syndrome (AIDS) has zero values for all patients thus we removed these attributes. Further, the missing values of attributes such as "number of pregnancies, hormonal contraceptive, and first sexual intercourse" are replaced with the help of mean imputation method which is expressed as:

$$X = \frac{1}{n} \left( \sum_{i=1}^{n} X_i \right) = \frac{X_1 + X_2 + \dots + X_n}{n}$$
(1)

However, each feature has different range of values due to which algorithms behave arbitrarily. To overcome this issue, we normalize all attributes in rang [0, 1] by using min-max normalization which can be expressed as:

$$v'_{i} = \frac{v_{i} - min_{A}}{max_{A} - min_{A}} (new \ max_{A} - new \ min_{A}) + new \ min_{A} \ (2)$$

where  $min_A$  and  $max_A$  represents the minimum and maximum values of the considered attribute, respectively. The complete dataset suffers from the issue of data imbalance because it has very less instances affected by the cancer therefore to address this issue; we apply Synthetic Minority Oversampling Technique (SMOTE) is used. The main idea of SMOTE is to perform oversampling to generate the synthetic samples rather than oversampling by replacements. The records belonging to the minority class are Isolated from the dataset. To generate synthetic samples, a single column is randomly selected from each record, and the difference between it and a neighbouring record is calculated. This process can be expressed as:

$$X_{syn} = X_i + rand(0,1) \times |X_{knn} - X_i|$$
(3)

where  $X_{knn}$  represents the nearest neighbour of x which is obtained by computing the Euclidean distance between  $X_i$  and every sample in the corresponding minority class set A.

## C. Proposed Deep Ensemble Model

This section presents the detailed discussion about proposed model where deep learning models are combined together to formulate the hybrid structure for pattern learning. The proposed model is based on the concept of CNN models where different layers are used to obtain the classification results. First of all, we present brief discussion about these layers.

# 1) Layer used in the Network

## a) Conv1D layer

The 1D convolutional layer applies convolutional filters to the input data. This layer is typically used to extract local features from the input by sliding filters over the input tensor and applying element-wise multiplication followed by summation. The first Conv1D layer has 64 filters, each with a kernel size of 3. Given an input tensor x with shape(n,d), where n is the sequence length and d is the input dimension, and a filter www of sizek, the output of the convolution operation at position i is:

$$y_i = \sum_{j=1}^{j} x_{i+j-1} \cdot w_j + b$$
 (4)

where *b* is the bias term. The ReLU activation function is then applied:

$$y_i^{ReLU} = max\left(0, y_i\right) \tag{5}$$

#### b) MaxPooling layer

The max-pooling layer reduces the dimensionality of the input by taking the maximum value over a sliding window of a specified size. In this case, the window size is 2. Max-pooling is used to down sample the feature map while retaining the most important features (i.e., those with the highest activation values). It can be expressed as:

$$y_i = \max(x_{i:i+p}) \tag{6}$$

where x is the tensor, and p is the window size.

#### *c) Flatten layer*

The flatten layer takes the multi-dimensional input and reshapes it into a 1D vector, which is required by the fully connected (Dense) layers. It does not perform any computation but simply transforms the data structure. According to this process, If the input feature map is of shape (m, n), where m is the number of filters and n is the length of each feature map, the output of the flatten layer is a vector of shape  $(m \times n)$ .

## d) Dense layer

The Dense layer is a fully connected layer that computes a weighted sum of its inputs, followed by an activation function. The purpose of this layer is to learn complex relationships between the features extracted by the previous layers. Given an input vector x of size n, the output of the Dense layer with mmm neurons is computed as:

$$y_j = \sum_{i=1}^n x_i \cdot w_{ij} + b_j \text{ for } j = 1, 2, 3.., m$$
(7)

where  $w_{ij}$  are the weights and  $b_j$  are the biases. The ReLU activation is applied as:

$$y_j^{ReLU} = max\left(0, y_j\right) \tag{8}$$

#### e) Dropout layer

The dropout layer is a regularization technique used to prevent overfitting by randomly setting a fraction of input units to 0 during training. In this model, 50% of the input units are dropped out. During training, each input unit  $x_i$  is either retained with probability p or set to zero with probability 1 - p:

$$x_{i}' = \begin{cases} \frac{x_{i}}{p} \text{ with probability } p\\ 0 \text{ with probability } 1 - p \end{cases}$$
(9)

## f) Output dense layer

The final Dense layer outputs a single value representing the probability of the input belonging to the positive class (cervical cancer present). The sigmoid activation function is used to squash the output to a value between 0 and 1. Given an input vector x, the output is computed as:

$$y = \sigma(\sum_{i=1}^{n} x_i \cdot w_i + b) \tag{10}$$

where  $\sigma(z)$  is the sigmoid activation function:

$$\sigma(z) = \frac{1}{1 + e^{-z}} \tag{11}$$

#### 2) Proposed hybrid architecture for classification

The ensemble model leverages multiple CNNs to improve the classification accuracy of cervical cancer detection. Each CNN is trained independently on the same dataset, and their predictions are averaged to form the final prediction. This technique, known as model averaging, aims to reduce the variance of predictions and increase overall robustness. The complete approach is carried out in following phases:

• Conv1D Layer: This layer extracts local patterns from the input sequence. In this work, this layer uses 64 filters, each of size 3, to perform convolution over the input data. The operations of this data can be expressed as:

$$y_i^{(1)} = ReLU(\sum_{j=1}^k w_j, x_{i+j-1} + b)$$
(12)

where  $y_i^{(1)}$  is the output, x is the input, w is the filter, k = 3 is the kernel size, and b is the bias term.

• Maxpooling and second conv1d Layer: this layer down samples the feature map by taking the maximum value in each pool of size 2. This reduces the dimensionality of the output and retains the most prominent features.

$$y_i^{(2)} = \max(y_{2i}^{(1)}, y_{2i+1}^{(1)})$$
(13)

With this layer, another convolution layer is applied to extract the features from the pooled output. This layer has 32 filters with the same kernel size of 3.

• Flatten Layer and dense layer: the flatten layer converts the 2D output from the previous layer into a 1D vector, preparing the data for the fully connected layers whereas the dense learns complex relationships in the data by applying a weighted sum and ReLU activation. It is expressed as:

$$z_{j} = \sum_{i=1}^{d} w_{ij} \cdot y_{i} + b_{j}$$
$$y_{i}^{(3)} = ReLU(z_{j})$$
(14)

• Dropout and dense output layer: the dropout layer prevents overfitting by randomly setting 50% of the input units to 0 during training whereas the dense layer is fully connected layer that outputs a probability score indicating the likelihood of the input belonging to the positive class. it can be expressed as:

$$z = \sum_{i=1}^{64} w_i y_i^{(4)} + b$$
$$y^{(5)} = \sigma(z)$$
(15)

Each CNN model (model1, model2, model3) is trained on the same dataset for 10 epochs with a batch size of 32. The training process adjusts the weights in each model to minimize the binary cross-entropy loss, using the Adam optimizer. After training, each model generates predictions on the test dataset, resulting in three prediction sets. The final prediction is computed as the average of these predictions:

$$final Prediction = \frac{pred1+pred2+pred3}{3}$$
(16)

This averaging process aims to reduce the variance and improve the stability of predictions. The final ensemble prediction is then threshold at 0.5 to produce binary class labels:

$$final \ prediction = \begin{cases} 1 \ if \ final \ Prediction > 0.5 \\ 0 \ , otherwise \end{cases} (17)$$

The complete architecture of CNN model is depicted in below given Fig. 2.



Fig. 2. Ensemble CNN model.

The proposed ensemble model combines the strengths of multiple individual models, reducing the likelihood of overfitting and improving generalization. By using this approach, the proposed model achieves a more robust and stable performance compared to single classifiers. In this study, the average prediction mechanism ensures that the final classification result benefits from the diversity of the individual models, minimizing the impact of any single model's bias or variance. Moreover, the cervical cancer detection often suffers from issues like class imbalance, noisy data, and feature irrelevance. The ensemble approach mitigates these challenges by blending the decision boundaries of multiple models, ensuring that outlier or minority class instances are not overlooked.

## IV. RESULT AND DISCUSSION

A This section presents the detailed outcome of proposed model and presents a comparative analysis to show the impact of proposed approach in cervical classification. The first subsection presents the description of dataset used in this work, next subsection presents the performance measurement parameter details, and finally, the comparative analysis is presented.

### A. Dataset Details

The dataset was sourced from the Hospital Universitario de Caracas in Caracas, Venezuela, and is publicly available as the Risk Factors dataset on the UCI (University of California, Irvine) Machine Learning repository. The dataset contains 36 attributes associated with cervical cancer risk, with four being categorical in nature. These categorical attributes reflect the results of various medical tests conducted to validate clinical findings related to cervical cancer. The Hinselmann test, also known as colposcopy, is used to assess whether lesions are cancerous. Schiller's test involves applying a solution to the targeted area to examine its potential malignancy. Cytology testing helps determine the presence of cancerous fluid in a specific area of the body. When other clinical tests are inconclusive, a complete biopsy is performed to provide a definitive diagnosis of cancer. Key risk factors in building a predictive model for cervical cancer include contraceptive pill usage, alcohol consumption, a high number of sexual partners, and other physiological parameters. In summary, the dataset captures information on lifestyle habits, such as smoking, sexual behaviour, and the results of medical tests. Notably, attributes like age, Number of Sexual Partners (NSP) and Hormonal Contraceptives (HC) exhibit a proper range of variation, while other attributes display deviations from their mean values, primarily due to their Boolean nature. The dataset also contains a significant number of missing values, necessitating the application of mean and median imputation methods to address these gaps. The

missing value imputation methodology is discussed in proposed model section. Table I demonstrates the attribute details of the cervical cancer.

Feature (F)	Description	Feature (F)	Description	Feature (F)	Description
F1	Age (Years)	F9	Hormonal Contraceptives	F17	STDs: Vulvo-perineal Condylomatosis
F2	Number of Sexual Partners	F10	IUD	F18	STDs: Syphilis
F3	First Sexual Intercourse (Age)	F11	IUD (Years)	F19	STDs: Pelvic Inflammatory Disease
F4	Number of Pregnancies	F12	STDs	F20	STDs: Genital Herpes
F5	Smokes	F13	STDs (Number)	F21	STDs: Molluscum Contagiosum
F6	Smokes (Years)	F14	STDs: Condylomatosis	F22	STDs: AIDS
F7	Smokes (Packs/Year)	F15	STDs: Cervical Condylomatosis	F23	STDs: HIV
F8	Hormonal Contraceptives	F16	STDs: Vaginal Condylomatosis	F24	STDs: Hepatitis B
F25	STDs: HPV	F26	STDs: Number of Diagnoses	F27	STDs: Time Since First Diagnosis
F28	STDs: Time Since Last Diagnosis	F29	Dx: Cancer	F30	Dx: CIN
F31	Dx: HPV	F32	Dx		

TABLEL	ATTRIBUTE D	ETAILS OF	THE CERVI	CAL CANCER
TIDEL I.	ATTRIDUTE D	LIAILSOI	THE CERVI	CAL CARCER

Hardware and software details: The proposed hybrid model for cervical cancer detection, which leverages deep learning techniques, is implemented using several widelyused libraries and frameworks that facilitate deep learning, data processing, and model training. It uses Tensorflow2.x, Keras, and CUDA as the main libraries for DL models. It uses Intel Core i7 (8th Gen) CPU with NVIDIA GeForce RTX 3060 GPU with 8GB of memory.

## B. Performance Measurement Parameters

This section describes the details of performance measurement parameters to evaluate the performance of classifier models. However, in medical domain, correctly diagnosing a patient with a disease is more crucial than simply confirming that someone is healthy. As a result, when evaluating a model, accuracy alone is insufficient. Therefore, we have considered six different performance metrics such as accuracy, sensitivity, specificity, precision, F-measure, and Area Under Curve (AUC). These performance metrics are obtained with the help of confusion matrix which consists of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). With the help of this confusion matrix, the other performance parameters can be computed. In this context, TP represents the correct identification of cancer in an individual who indeed has the disease, while TN denotes the correct prediction of a non-cancerous person as cancer-free. FP, or false positive, occurs when a person without cancer is incorrectly identified as having the disease, and FN, or false negative, refers to a cancerous individual being wrongly classified as cancer-free. Table II below describes the parameter, its definition and expression to compute the performance.

Metric	Definition	Formula
Accuracy	Proportion of correct predictions out of the total predictions made by the model.	$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$
Sensitivity	Ability of the model to correctly identify individuals with cervical cancer (also known as Recall).	$Sensitivity = \frac{TP}{TP + FN}$
Specificity	Ability of the model to correctly identify individuals without cervical cancer.	$Specificity = \frac{TN}{FP + TN}$
Precision	Proportion of individuals predicted to have cervical cancer who were correctly identified.	$Precision = \frac{TP}{TP + FP}$
F-Measure	Harmonic mean of precision and sensitivity, balancing the model's performance.	$FMeasure = \frac{2 \times TP}{2 \times TP + FP + FN}$
ROC Curve	The Receiver Operating Characteristic (ROC) curve is a graphical representation used to visualize the model's performance across different classification thresholds. It demonstrates the trade-off between sensitivity and specificity.	
AUC	AUC measures the overall area under the ROC curve, providing a single scalar value that indicates the model's performance. A larger AUC corresponds to a better-performing model.	

TABLE II. METRIC PARAMETER, ITS DEFINITION AND EXPRESSION TO COMPUTE THE PERFORMANCE

#### C. Comparative Analysis

In this subsection, we present a comparative analysis where performance of proposed approach is compared with existing classification models. The overall confusion matrix using proposed model is demonstrated in below given Fig. 3 as confusion matrix. Lilhore *et al.* [31] have presented a study on cervical cancer classification by using machine learning methods such as random forest, SVM and decision tree method. Table III shows the comparative performance for random forest classifier.

According to this experiment, the class-level performance shows that for class-0, precision is reported

as 0.91 which shows that 91% instances were correctly predicted corresponding to class-0 test data whereas recall is obtained as 94% and finally, F1-Score is reported as 93% which shows the harmonic mean of precision and recall for class-0. The AUC for this experiment is reported as 0.555, despite of achieving high precision, recall, and F1-Scores for individual classes, the low AUC suggests the model struggles with overall separability between the classes, possibly due to issues like class imbalance, overfitting, or insufficient training data. The final accuracy of this model is reported as 95%.



TABLE III. CLASSIFICATION REPORT OF RANDOM FOREST CLASSIFIER

Class	Precision	Recall	F1-Score
0	0.91	0.94	0.93
1	0.78	0.9	0.88
Macro Avg.	0.889	0.875	0.757
Weighted Avg.	0.96	0.95	0.96

Gnanavel *et al.* [20] reported the classification report of SVM classifier method and measured the performance in terms of precision, recall, F1-Score and accuracy. Table IV given below shows the performance analysis of SVM classifier.

TABLE IV. CLASSIFICATION REPORT OF SVM CLASSIFIER

Class	Precision	Recall	F1-Score
0	0.901	0.87	0.89
1	0.735	0.798	0.875
Macro Avg.	0.845	0.812	0.684
Weighted Avg.	0.912	0.875	0.842

According to this experiment, the SVM classifier has reported the overall accuracy of 91% which is lower than the aforementioned Random Forest Classifier. The precision, recall and F1-Score for class-0 are reported as 0.901, 0.87, and 0.89 whereas the performance in terms of these parameters for class-1 is reported as 0.735, 0.798, and 0.875, respectively. The overall, AUC is obtained as 0.510 which shows the poor performance to spate the classes for the complete experiment.

The next experiment presents the outcome of decision tree method where performance of classifier is measured in terms of precision, recall and F1-Score. Table V demonstrates the obtained performance.

TABLE V. PERFORMANCE ANALYSIS FOR DECISION TREE CLASSIFICATION

Class	Precision	Recall	F1-Score
0	0.91	0.94	0.93
1	0.78	0.90	0.88
Macro Avg.	0.889	0.875	0.757
Weighted Avg.	0.96	0.95	0.96

In this experiment, the average accuracy is reported as 95% whereas the overall AUC performance is reported as 0.555. The precision, recall and F1-Score for class-0 are reported as 0.9, 0.94, and 0.93, respectively whereas performance in terms of these parameters is reported as 0.78, 0.90 and 0.88, respectively. Finally, we present the outcome of proposed classifier model and measured its performance in terms of aforementioned parameters. Table VI shows the obtained performance by using proposed model.

TABLE VI. CLASSIFICATION REPORT OF PROPOSED DEEP ENSEMBLE CLASSIFIER

Class	Precision	Recall	F1-Score
0	0.98	0.98	0.98
1	0.57	0.50	0.53
Macro Avg.	0.77	0.74	0.76
Weighted Avg.	0.96	0.96	0.96

For this experiment, the average classification accuracy is obtained as 96% and AUC performance is reported as 0.99 which shows a significant improvement in the overall performance of the proposed classification model. The performance of proposed model in terms of precision, recall and F1-Score for class -0 is reported as 0.98, 0.98 and 0.98 whereas for class -1, 0.57, 0.50, and 0.53, respectively. Finally, we present a comparative analysis where performance of proposed model is compared with existing models for cervical cancer classification. Below given Fig. 4 depicts the comparative analysis of Random Forest, SVM, Decision tree and proposed model.



Further, the outcome of this approach is compared with state-of-art methods. Table VII shows this comparative analysis for the considered dataset. The existing methods are described in [32].

TABLE VII. COMPARATIVE PERFORMANCE ANALYSIS

Algorithm

Precision Recall F1-Score Accura

Algonithm	1 I CUSION	Recall	1-30016	Accuracy
Algorithm	(%)	(%)	(%)	(%)
AdaBoost	97.53	96.34	96.93	94.19
XGBoost	96.91	97.52	97.21	94.77
Multi-Layer Perceptron	96.3	98.73	97.5	95.35
Support Vector Machine	95.68	99.36	97.48	95.35
Gradient Boosting Machine	96.91	99.37	98.13	96.51
Logistic Regression	98.77	98.16	98.46	97.09
Cat Boost	98.15	98.76	98.45	97.07
Decision Tree	96.3	98.11	97.2	94.77
Random Forest	98.15	96.36	97.25	94.77
Naïve Bayes (GaussianNB)	87.65	99.3	93.11	87.79
Ensemble Classifier+SMOTE	96.89	96.3	96.59	96.58
Proposed Model	98.9	99.5	99.1	97.5

This dataset has been studied widely in various data mining related tasks. Table VIII below presents a comparative analysis where performance of proposed model is compared with the state-of-art ML models discussed in [33].

TABLE VIII. COMPARATIVE ANALYSIS WITH EXISTING METHODS

Machine Learning Models	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Random Forest	96%	96%	96%	96%
AdaBoost	95%	95%	95%	95%
Gradient Boosting	97%	98%	97%	97%
MLP	96%	96%	96%	96%
XGB	98%	98%	98%	98%
Decision Tree	96%	98%	96%	97%
Logistic Regression	95%	74%	81%	77%
SVM	96%	96%	96%	96%
Gaussian NB	89%	96%	89%	91%
Proposed Model	97.5	98.9	99.5	99.1

According to this experiment, the proposed model has reported the highest accuracy as 96% with the highest AUC of 0.99 (Fig. 5) which shows a significant improvement in the overall performance.



Fig. 5. ROC curve analysis.

Figs. 6 and 7 depict the performance analysis for 50 epochs where training and validation accuracy for CNN models is depicted in Figs. 4 and 5 show the training and validation loss.

The outcome of proposed approach and its comparative analysis helps to address the research questions thoroughly and this outcome is used to derive insights.



Fig. 6. Training and validation accuracy.



Fig. 7. Training and validation loss.

**Solution for RQ 1:** The comparative performance of the proposed model against Random Forest, SVM, and Decision Tree is summarized in Tables IV–VII where it shows that the proposed deep ensemble classifier has reported the highest accuracy of 96% whereas Random Forest, SVM and Decision Tree have reported the accuracy as 95%, 91% and 95%, respectively. This shows that the proposed approach outperformed existing classification models. For class-0, the proposed model achieves the highest precision, recall, and F1-Score (0.98 each), significantly outperforming all traditional methods.

For class-1, the performance of the proposed model is lower in terms of precision (0.57), recall (0.50), and F1-Score (0.53). The AUC for the proposed model (0.99) is markedly higher than that of Random Forest (0.555), SVM (0.510), and Decision Tree (0.555), highlighting its ability to distinguish between classes effectively.

Solution for RQ 2: The proposed model has reported high AUC, precision, recall when compared with the existing methods. However, the proposed method faces challenges while handling the class imbalance and limited data.

**Solution for RQ 3:** The AUC analysis represents the substantial class separability performance which indicates that the proposed deep ensemble model effectively differentiates between the two classes, reducing false positives and false negatives, even though the class-1 metrics are suboptimal.

**Solution for RQ 4:** The proposed model leverages the strengths of multiple classifiers, ensuring robust decision-making by combining predictions. Moreover, the proposed approach incorporates deep learning approach which helps to capture more intricate patterns in the data, resulting in improved learning.

**Solution for RQ 5:** the imbalanced data affects the performance of this approach because it causes biasness towards the high sample data therefore, advanced data augmentation or increasing the data samples are the possible solution. Moreover, pre-trained deep learning models can be incorporated as feature extraction for minority samples.

Limitations of proposed model and possible solutions to address these issues: as discussed, the proposed model is based on the combination of multiple CNN which may suffer from overfitting for limited training data, moreover, insufficient regularization also can lead to the overfitting issue. These issues can be addressed by increasing dropout rate, early stopping, data augmentation and L2 regularization. On the other hand, the underfitting occurs when the model is too simple or lacks the capacity to capture the underlying patterns in the data however, this model uses combination of three CNN architecture which is cable in capturing the patterns.

## V. CONCLUSION

Cervical cancer is the most prevalent cancer among women and the most common surgical procedures to treat this cancer is hysterectomy. Its screening process requires significant time and resources. In underdeveloped countries, conducting these screenings is often not feasible due to limited medical facilities, including a shortage of physicians and equipment. Moreover, the early detection of cancer can play significant role to reduce the mortality rate. Thus, ML based automated processes are widely adopted in these applications for early detection of diseases however the ML based solutions suffer from several challenging issues such as poor accuracy, computational complexity, and inefficiency in handling the large size dataset. To overcome these issues, we have introduced a deep learning-based framework where CNN based models are combined together to train the model and an averaging method is applied to obtain the final prediction. The proposed approach is validated using publicly available dataset and experimental analysis shows that the proposed approach reported highest classification accuracy. In future work, the deep learning-based classification frameworks can be extended to image data to further improve the cervical cancer detection accuracy. Moreover, the data augmentation methods can be incorporated to address the overfitting issues.

#### CONFLICT OF INTEREST

The authors declare no conflict of interest.

#### AUTHOR CONTRIBUTIONS

Aishwarya N. Kumar has done the abstract, introduction, literature review, data collection and simulation of results along with conclusion. Meenakshi Sundaram A. has reviewed the paper advised corrections. Both authors had approved the final version.

#### REFERENCES

- P. Nenclares and K. J. Harrington, "The biology of cancer," *Medicine*, vol. 48, no. 2, pp. 67–72, 2020.
- [2] E. Schuster, R. Taftaf, C. Reduzzi, M. K. Albert, I. Romero-Calvo, and H. Liu, "Better together: Circulating tumor cell clustering in metastatic cancer," *Trends in Cancer*, vol. 7, no. 11, pp. 1020–1032, 2021.
- [3] A. Bansal, S. Goyal, A. Goyal, and M. Jana, "WHO classification of soft tissue tumours 2020: An update and simplified approach for radiologists," *European Journal of Radiology*, vol. 143, 109937, 2021.
- [4] P. Bisoyi, "Malignant tumors—As cancer," in *Understanding Cancer*, Academic Press, 2022, pp. 21–36.
- [5] J. Boutry, S. Tissot, B. Ujvari, J. P. Capp, M. Giraudeau, A. M. Nedelcu, and F. Thomas, "The evolution and ecology of benign tumors," *Biochimica et Biophysica Acta (BBA)*—*Reviews on Cancer*, vol. 1877, no. 1, 188643, 2022.
- [6] J. Ferlay, M. Colombet, I. Soerjomataram, C. Mathers, D. Parkin, M. Piñeros, A. Znaor, and F. Bray, "Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods," *International Journal of Cancer*, vol. 144, no, 8, pp. 1941–1953, 2019.
- [7] F. Islami, L. A. Torre, J. M. Drope, E. M. Ward, and A. Jemal, "Global cancer in women: Cancer control priorities," *Cancer Epidemiology, Biomarkers & Prevention*, vol. 26, no. 4, pp. 458–470, 2017.
- [8] S. Shinde, N. K. Vishvakarma, A. K. Tiwari, V. Dixit, S. Saxena, and D. Shukla, "Therapeutic options for the management of cervical cancer," in *A Theranostic and Precision Medicine Approach for Female-Specific Cancers*, Academic Press, 2021, pp. 193–212.
- [9] K. S. Tewari and B. J. Monk, "Evidence-based treatment paradigms for management of invasive cervical carcinoma," *Journal of Clinical Oncology*, vol. 37, no. 27, pp. 2472–2489, 2019.
- [10] R. A. Smith, K. S. Andrews, D. Brooks, S. A. Fedewa, D. Manassaram-Baptiste, D. Saslow, and R. C. Wender, "CA: A cancer journal for clinicians 69, 184," *CA: A Cancer Journal for Clinicians*, vol. 69, 184, 2019.
- [11] Monica and R. Mishra, "An epidemiological study of cervical and breast screening in India: District-level analysis," *BMC Women's Health*, vol. 20, pp. 1–15, 2020.
- [12] World Health Organization (WHO). (2018). India factsheet. International Agency for Research on Cancer. [Online]. Available: https://gco.iarc.fr/today/data/factsheets/populations/356-india-factsheets.pdf
- [13] T. Ramamoorthy, V. Kulothungan, K. Sathishkumar, N. Tomy, R. Mohan, S. Balan, and P. Mathur, "Burden of cervical cancer in India: estimates of years of life lost, years lived with disability and disability adjusted life years at national and subnational levels using the National Cancer Registry Programme data," *Reproductive Health*, vol. 21, no. 1, p. 111, 2024.
- [14] N. Shanker, P. Mathur, P. Das, K. Sathishkumar, A. M. Shalini, and M. Chaturvedi, "Cancer scenario in North-East India & need for an appropriate research agenda," *Indian Journal of Medical Research*, vol. 154, no. 1, pp. 27–35, 2021.
- [15] C. Taliento, G. Scutiero, M. Arcieri, G. Pellecchia, V. Tius, G. Bogani, and G. Vizzielli, "Simple hysterectomy versus radical hysterectomy in early-stage cervical cancer: A systematic review and meta-analysis," *European Journal of Surgical Oncology*, 108252, 2024.

- [16] K. S. Okunade, "Human papillomavirus and cervical cancer," *Journal of Obstetrics and Gynaecology*, vol. 40, no. 5, pp. 602–608, 2020.
- [17] P. L. Sachan, M. Singh, M. L. Patel, and R. Sachan, "A study on cervical cancer screening using pap smear test and clinical correlation," *Asia-Pacific Journal of Oncology Nursing*, vol. 5, no. 3, pp. 337–341, 2018.
- [18] W. Wu and H. Zhou, "Data-driven diagnosis of cervical cancer with support vector machine-based approaches," *IEEE Access*, vol. 5, pp. 25189–25195, 2017.
- [19] H. Alquran, W. A. Mustafa, I. A. Qasmieh, Y. M. Yacob, M. Alsalatie, Y. Al-Issa, and A. M. Alqudah, "Cervical cancer classification using combined machine learning and deep learning approach," *Computers, Materials & Continua*, vol. 72, no. 3, pp. 5117–5134, 2022.
- [20] N. Gnanavel, P. Inparaj, N. Sritharan, D. Meedeniya, and P. Yogarajah, "Interpretable cervical cell classification: A comparative analysis," in *Proc. 2024 4th International Conference* on Advanced Research in Computing (ICARC), Belihuloya, Sri Lanka, 2024, pp. 7–12. doi: 10.1109/ICARC61713.2024.10499737
- [21] G. Kumawat, S. K. Vishwakarma, P. Chakrabarti, P. Chittora, T. Chakrabarti, and J. C. W. Lin, "Prognosis of cervical cancer disease by applying machine learning techniques," *Journal of Circuits, Systems and Computers*, vol. 32, no. 01, 2350019, 2023.
- [22] M. M. Ali, K. Ahmed, F. M. Bui, B. K. Paul, S. M. Ibrahim, J. M. Quinn, and M. A. Moni, "Machine learning-based statistical analysis for early-stage detection of cervical cancer," *Computers in Biology and Medicine*, vol. 139, 104985, 2021.
- [23] L. W. Habtemariam, E. T. Zewde, and G. L. Simegn, "Cervix type and cervical cancer classification system using deep learning techniques," *Medical Devices: Evidence and Research*, pp. 163– 176, 2022.
- [24] A. Tripathi, A. Arora, and A. Bhan, "Classification of cervical cancer using deep learning algorithm," in *Proc. 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, May 2021, pp. 1210–1218.
- [25] H. Chen, J. Liu, Q. M. Wen, Z. Q. Zuo, J. S. Liu, J. Feng, and D. Xiao, "CytoBrain: Cervical cancer screening system based on deep

learning technology," Journal of Computer Science and Technology, vol. 36, pp. 347–360, 2021.

- [26] V. Chandran, M. G. Sumithra, A. Karthick, T. George, M. Deivakani, B. Elakkiya, and S. Manoharan, "Diagnosis of cervical cancer based on ensemble deep learning network using colposcopy images," *BioMed Research International*, vol. 2021, no. 1, 5584004, 2021.
- [27] M. Kalbhor, S. Shinde, D. E. Popescu, and D. J. Hemanth, "Hybridization of deep learning pre-trained models with machine learning classifiers and fuzzy min-max neural network for cervical cancer diagnosis," *Diagnostics*, vol. 13, no. 7, p. 1363, 2023.
- [28] Z. Alyafeai and L. Ghouti, "A fully-automated deep learning pipeline for cervical cancer classification," *Expert Systems with Applications*, vol. 141, 112951, 2020.
- [29] S. Jahan, M. S. Islam, L. Islam, T. Y. Rashme, A. A. Prova, B. K. Paul, and M. K. Mosharof, "Automated invasive cervical cancer disease detection at early stage through suitable machine learning model," *SN Applied Sciences*, vol. 3, pp. 1–17, 2021.
- [30] J. Lu, E. Song, A. Ghoneim, and M. Alrashoud, "Machine learning for assisting cervical cancer diagnosis: An ensemble approach," *Future Generation Computer Systems*, vol. 106, pp. 199–205, 2020.
- [31] U. K. Lilhore, M. Poongodi, A. Kaur, S. Simaiya, A. D. Algarni, H. Elmannai, and M. Hamdi, "Hybrid model for detection of cervical cancer using causal analysis and machine learning techniques," *Computational and Mathematical Methods in Medicine*, vol. 2022, no. 1, 4688327, 2022.
- [32] M. S. Ali, M. M. Hossain, M. A. Kona, K. R. Nowrin, and M. K. Islam, "An ensemble classification approach for cervical cancer prediction using behavioral risk factors," *Healthcare Analytics*, vol. 5, 100324, 2024.
- [33] M. Hasan, J. Islam, M. Al Mamun, A. A. Mim, S. Sultana, and M. S. H. Sabuj, "Optimizing cervical cancer prediction, harnessing the power of machine learning for early diagnosis," in *Proc. 2024 IEEE World AI IoT Congress (AIIoT)*, May 2024, pp. 552–556.

Copyright © 2025 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (<u>CC BY 4.0</u>).