IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 14, No. 2, April 2025, pp. 1450~1460
ISSN: 2252-8938, DOI: 10.11591/ijai.v14.i2.pp1450-1460

Deep ensemble learning with uncertainty aware prediction ranking for cervical cancer detection using Pap smear images

Sreelatha[1,2], Vrinda Shivashetty[3]
[1]Department of Information Science and Engineering, Sai Vidya Institute of Technology, Visvesvaraya Technological University, Belagavi, India
[2]Department of Computer Science and Engineering, Presidency University, Bangalore, India
[3]Department of Information Science and Engineering, Sai Vidya Institute of Technology, Bangalore, India

Article Info

Article history:

Received Mar 30, 2024
Revised Nov 5, 2024
Accepted Nov 14, 2024

Keywords:

Cervical cancer
Deep learning
Ensemble learning
Pap smear
Uncertainty aware prediction ranking

ABSTRACT

This paper proposes a novel deep ensemble learning framework designed for the efficient detection and classification of cervical cancer from Pap smear images. The proposed study implements three advanced learning models namely DenseNet201, Xception, and a classical convolutional neural network (CNN) customized with optimal hyperparameters to automate feature extraction and cervical cancer detection process. The proposed study also introduces a novel ensemble learning to enhance the classification of cervical cancer. The proposed ensemble mechanism is based on the confidence aggregation followed by uncertainty quantification and prediction ranking scheme, thus ensuring that more reliable predictions have a proportionally greater influence on the final outcome. The primary goal is to leverage the collective intelligence of the ensemble in a manner that prioritizes reliability and minimizes the impact of less certain predictions. The experimental analysis is carried out on two dataset one with whole slide images (WSI) and another on cropped images. The proposed ensemble model achieves an accuracy rate 100 and 97% for dataset with WSI and with cropped images respectively.

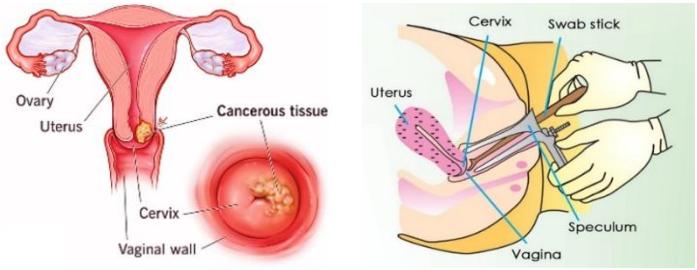
Corresponding Author:

Sreelatha
Department of Information Science and Engineering, Sai Vidya Institute of Technology
Visvesvaraya Technological University
Belagavi, India
Email: sreelatha.pk@presidencyuniversity.in

1. INTRODUCTION

Digitization and recent advances in the healthcare system have made treatment planning more efficient and enhanced medical care [1]. Modern healthcare systems, therefore, recognize the critical role of early detection and diagnosis, especially in the fight against cancer, which is one of the leading causes of morbidity and mortality worldwide [2]. Moreover, cancer is a multifactorial disease and remains incurable when it reaches its late stages [3]. Therefore, because of this complexity, an early diagnosis is important for effective treatment and management. Among many cancers, Figure 1 shows an illustrative overview of cervical cancer and the Pap smear procedure. Cervical cancer shown in Figure 1(a) remains a significant health problem for women and according to the World Health Organization [4], [5], it is the fourth most common cancer among women worldwide. Although cervical cancer is one of the most preventable and treatable cancers, it is often less known than other cancer types due to several reasons, such as lack of routine screening, public awareness, and socio-cultural issues [6].

There are many works in the existing literature discussing various screening techniques for detection and diagnosis of cervical cancer. These screening techniques include Pap smear test, colposcopy, visual inspection with acetic acid (VIA), human papillomavirus (HPV) test, and biopsy. However, each of these screening techniques has its own advantages and disadvantages [7]. HPV based screen test is mostly adopted to detect high-risk strains of the virus, so it is not so good in detecting cancer directly. If a positive result is obtained, it requires additional testing. On the other hand, VIA based screening is considered a cost-effective method which involves applying vinegar to the cervix to highlight abnormal cells [8]. However, it often produces false-positive results and it also requires highly trained medical professionals. The other screening techniques, such as colposcopy and biopsy, provide a detailed assessment using magnification tools and tissue analysis but are invasive and painful procedures [9]. However, the Pap smear test, as seen in Figure 1(b), is frequently used because of its ease of use, low level of discomfort, and screening effectiveness. In order to distinguish between precancerous and cancerous cells it entails removing cells from the cervix for microscopic analysis.



(a) (b)

Figure 1. Illustrative overview of cervical cancer and the Pap smear procedure of (a) cancerous tissue within the cervix and (b) cell collection from the cervix

Figure 1(a) provides a schematic illustration of the anatomy of the female reproductive system with localization of cervical cancerous tissue within the cervix located between the uterus and the vaginal wall. Another part in Figure 1(b) shows the scenario of a Pap smear-based screening procedure, where cells are collected from the cervix using a swab stick during a gynecological examination. The collected cells are then smeared on a microscopic slide and stained for better visualization under the microscope. Basically, a Pap smear image is a high-resolution data taken from this prepared slide that can be used to analyze individual cervical cells and their characteristics, such as size, shape, and nucleus [10]. However, the accuracy of the Pap smear test is computed by cytotechnologists and pathologists, which is often a time-consuming process, subjected to human error and subjective variability. The Pap smear images often associated with varying illumination, artifacts, and cells in Pap smear images often overlap or form clusters, which complicates the process of accurate identification of individual cells. Another issue is the variability in cell morphology, intra-class variability and inter-class similarity, making it challenging to extract meaningful features that can effectively distinguish between cell types. Therefore, detecting a cell as cancerous or non-cancerous using Pap smear images requires a sophisticated image processing and complex feature engineering methods. The use of artificial intelligence (AI) in healthcare has shown promising developments in the area of diagnostic imaging [11]. The machine learning models have emerged as a robust tool for analyzing complex medical images. This not only reduces time consumption in the manual examination but also automates the screening process, reduces the rate of misdiagnosis, and standardizes the detection process, thereby ensuring timely intervention [12]. However, one biggest problem here is an achieving optimization in the feature generalization and learning process. Apart from this handling overfitting and underfitting issues often associated deep learning (DL) model is quite tedious task.

In the recent state-of-art-works, many researchers have suggested data driven and predictive schemes to identify and categorize cervical cancer. Palanisamy *et al.* [13] combined DL model with wavelet transform for identification and classification of cervical cancer using Pap smear images. Kalbhor *et al.* [14] explored the use of pre-trained DL models for feature extraction from images, and used logistic regression algorithms for classification task. Habtemariam *et al.* [15] introduced an integrated system for cervix type and cervical cancer classification using deep efficient models for feature extraction and classification. Yu *et al.* [16] considered convolutional neural network (CNN) model and assess its performance by combining it with spatial pyramid pooling layers and inception modules to perform cervical cancer classification task. Ghoneim *et al.* [17] implemented extreme learning machines (ELM) and leveraging transfer learning from

CNN. The presented predictive model achieves good accuracy rates. Cheng *et al.* [18] presented a classification model by integrating low- and high-resolution whole slide images (WSI) with recurrent neural network (RNN) to detect and classify lesion cells. Alsubai *et al.* [19] used CNN with four hidden layers for distinguish cervical cells into healthy, precancerous, and benign cervical cells. Chauhan *et al*. [20] suggested a hybrid DL model with concatenated features from VGG-16, ResNet-152, and DenseNet-169 to develop predictive model for both binary and multi-class classification problem. Alsalatie *et al.* [21] introduced an ensemble DL model for the automatic diagnosis of cervical cancer by focusing on the WSI image rather than individual cells. Another study towards ensemble learning is conducted by Ilyas and Ahmad [22]. This study proposes a majority voting scheme to enhance the accuracy of cervical cancer diagnosis by integrating different classifiers such as decision tree, support vector machine (SVM), and random forest. Chadaga *et al.* [23] designed stacked ensemble learning scheme to predict cervical cancer risk followed by statistical techniques for feature selection, and the Borderline-SMOTE method to addressed data imbalance issues. Gupta and Gupta [24] applied feature selection and data balancing techniques to enhance the feature generalization capability for DL models. The authors have then performed ensemble learning to predict cervical cancer. Deo *et al*. [25] introduced an advanced learning scheme that utilizes a transformer model with attention mechanisms for classifying cervical cancer from Pap smear images. Mahajan and Kaur [26] considered an application of ResNet-50 in cervical cancer cell classification task. However, this work also focuses on preprocessing, input image segmentation using *k*-means clustering. The segmented cell images are then introduced to ResNet-50 for the classification task. Table 1 presents summary of the above-mentioned literature to offer quick insight to understand the current research trend.

Table 1. Summary of the literature on cervical cancer classification using Pap smear images and DL

Citations	Research objective	Method adopted	Dataset/Result	Remark
[13]	Multi-class classification.	Wavelet transform, ResNet-18 and data augmentation	Open access dataset; 99% average detection index.	Generalizability not assessed
[14]	Predict cervical cancer from Pap smear images.	Alexnet, Resnet-18, Resnet-50, Googlenet for feature extraction and logistic regression for classification.	Dataset name not specified, 95.14% with AlexNet.	Lacks extensive discussion on model interpretability and clinical implications of false positives/negatives.
[15]	Develop an automatic system for cervical cancer classification.	MobileNetv2-YOLOv3 for ROI extraction, and EfficientNetB0 for classification.	Colposcopy and histopathology images, achieved 99.88% mAP	Limited adaptability to new techniques.
[16]	Distinguish abnormal from normal cells.	CNN + inception + pyramid pooling	ThinPrep cytologic test dataset, Achieved AUC of 0.997.	Computational efficiency not discussed.
[17]	Detect and classify cervical cancer cells.	CNNs for feature extraction and an ELM classifier for classification	Herlev database; 99.5% accuracy for detection and 91.2% for classification.	ELM's comparative advantage unclear.
[18]	Enhance cervical cancer screening on WSIs.	RNN-based WSI classification model.	Dataset name not specified, 93.5% Specificity, 88.5% true positive rate for 10 lesion cells.	High complexity and computational demands.
[19]	Classify cervical cells into five categories.	Deep CNN with four convolutional layers	SIPaKMeD dataset; 91.13% accuracy.	Needs optimization in computational requirement
[20]	Detect cervical cancer via binary and multiclass classification.	Hybrid Deep Feature Concatenated Network with two-step data augmentation.	SIPaKMeD database; 97.45% for 5-class and 99.29% for 2-class.	No effective benchmarking
[21]	Automate diagnosis of WSI for cervical cancer.	Full ensemble DL model	Dataset name not specified; 99.6% accuracy.	Focuses on simplicity, butno effective benchmarking
[22]	Enhance diagnosis accuracy with ensemble classification.	Ensemble classification based on majority voting from a range of classifiers.	Dataset name not specified; 94% accuracy.	Addressed biases and overfitting, may face scalability issue
[23]	Screen for cervical cancer risk with ML.	Custom stacked ensemble learning, statistical methods for feature selection.	98% accuracy, 97% precision, 99% recall, 98% F1-score, 100% AUC.	Uses explainable AI for clinician transparency
[24]	Automate cervical cancer diagnosis.	Ensemble learning with feature selection and data balancing on the UCI dataset.	99.7% AUC with the cervical cancer dataset from the University of California Irvine database.	Showed computational efficiency with high accuracy, needs more evaluation
[25]	Classify cervical cancer from pap images.	Cross-attention-based transformer approach, utilizing the CerviFormer model.	96.67% Accuracy on SIPaKMeD data; 94.57% on Herlev data.	Highly complex model and large training cost
[26]	Categorize cervical cancer cells with DL.	Preprocessing, segmentation, and ResNet-50 for classification	97.4% accuracy, approx. 98% kappa score on Pap smear tests.	Reduces false-positive rates in manual screenings, but not fully automated

From the literature review, it can be seen that in the context of cervical cancer detection, many research works have been proposed to improve the diagnostic accuracy of Pap smear images using different DL models. However, significant issues remain that hinder the widespread adoption and effectiveness of these existing solutions. Potential issues are highlighted as follows: i) hospitals often use different Pap smear capture technologies, resulting in differences in image quality, color balance, and overall appearance. Existing models trained on specific datasets may struggle to cope with this variability, limiting generalizability across different populations and healthcare settings; ii) according to the analysis, most existing methods focus on achieving higher accuracy rather than cost optimization in data preparation tasks, feature engineering, and training models; iii) existing solutions have limited generalizability, and have high computational bottlenecks. There are still concerns about the computational efficiency and scalability of existing DL models, especially when deployed in resource-limited environments; and iv) another problem is that many current methods lack sufficient discussion of their mechanisms to justify their findings and model scope for applicability to the healthcare sector.

Therefore, this paper aims to offer a highly automated and responsive computational model that addresses the critical needs for adaptability, efficiency, and responsive in cervical cancer detection. The proposed model implements advanced DL architectures to automate Pap smear image analysis to reduce reliance on manual interpretation and potential subjectivity. The study also introduces a novel deep ensemble learning mechanism based on the confidence aggregation followed by uncertainty quantification and prediction ranking scheme, thus ensuring that more reliable predictions have a proportionally greater influence on the final outcome. The primary goal is to leverage the collective intelligence of the ensemble in a manner that prioritizes reliability and minimizes the impact of less certain predictions. The next section details the system architecture and proposed ensemble learning technique.

2. METHOD

The prime aim of the research work reported in this paper is to design highly adaptive and comprehensive computational model leveraging advanced learning strategy to effectively enhance the identification and classification of cervical cancer using Pap smear datasets. The study also intended to make model fully automated by minimizing the manual effort typically associated with data preparation without compromising the robustness and accuracy of the models used in the identification process. This process involves simplified preprocessing operations that ensures the consistency and uniformity within the dataset. This automation enables the system to prepare data rapidly and efficiently for model training. The schematic architecture followed by workflow overview is presented in Figure 2.

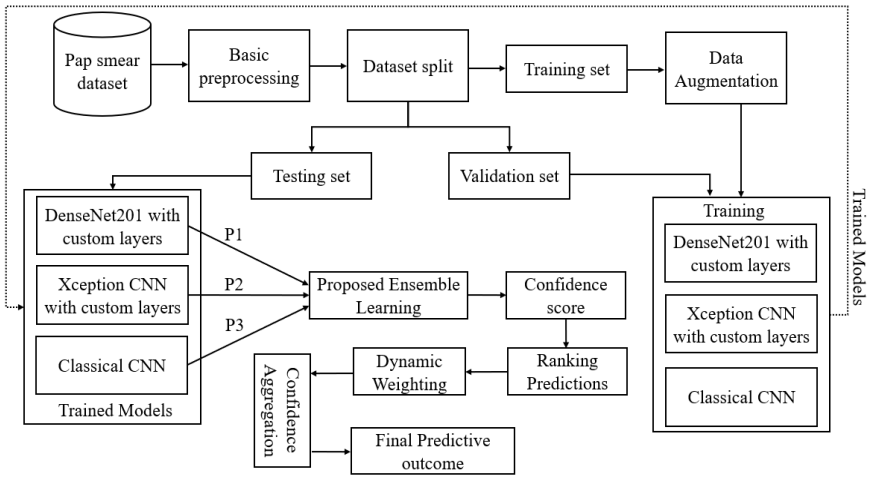


Figure 2. Illustrating system design and workflow overview of the proposed computational scheme for cervical cancer identification and classification

As can be seen from the schematic workflow shown in Figure 2, the proposed system integrates various computational operations in a highly synchronized manner, such as the initial data preparation stage, including basic preprocessing, data segmentation, and data augmentation. It then consists of DL architectures including DenseNet, Xception, and classic CNN models. Furthermore, a novel ensemble learning technique

is proposed to obtain final prediction results on Pap smear images based on the aggregation of quantified reliable confidence scores for each model. The next subsection discusses each component of the proposed system in detail.

2.1. Initial data processing

This phase of the proposed computational model intended to make model fully automated by minimizing the manual effort typically associated with data preparation before training the predictive models. The computing steps adopted for implementing data processing operation is discussed in Algorithm 1.

Algorithm 1. Automated data preparation and augmentation
Inputs: $D=\{xi|i=1, ..., N\}$ a set of N raw Pap smear images, d(target dimension for image resizing), α (proportion of the dataset to be used for training) and β(proportion of the dataset to be used for validation)
Output: D_{train}, D_{val}, D_{test} (preprocessed and partitioned datasets), and D_{aug} (augmented training set)
Start

1. Initialize $D'=\{\}$ as an empty vector to store preprocessed images.
2. For each image x_i in the raw dataset D:
   $x_i' \leftarrow T(R(xi,d))$, where $R:\mathrm{R}^{wi\times hi}\rightarrow\mathrm{R}^{d\times d}$ is a resizing function mapping image x_i from its original width w_i and height h_i to the target dimension d, and T is a transformation function that performs standard preprocessing steps on x_i.
3. Append the processed image x_i' to the list D'.
4. Partition the preprocessed dataset D' into D_{train}, D_{val}, D_{test} using function π:
   $(D_{train}, D_{val}, D_{test}) \leftarrow \pi(D',\alpha,\beta)$,
   where π is defined as: $\pi(D',\alpha,\beta)=\{D'_{[1,N\alpha]}, D'_{(N\alpha,N\alpha+N\beta]}, D'_{(N\alpha+N\beta,N]}\}$, with $N_\alpha=\lfloor\alpha\cdot N\rfloor$ and $N_\beta=\lfloor\beta\cdot N\rfloor$.
5. Initialize $D_{aug}=\{\}$ as an empty vector for augmented images.
6. For each image x_j in D_{train}:
   For k augmentations:
   $x_{aug} \leftarrow A(x_j)$ where A is an augmentation function that applies a set of stochastic transformations to x_j, such as random rotations R_θ, scaling S_λ, and flipping F.
7. Append the augmented image x_{aug} to D_{aug}.
8. Repeat steps 6 and 7 for k iterations to generate multiple augmented versions of each image in D_{train}.
9. Return D_{train}, D_{val}, D_{test}, D_{aug}.

End

The steps of Algorithm 1 are implemented to perform initial data preprocessing, which considers several inputs. After successful execution, it returns an output consisting of the preprocessed and partitioned datasets and the augmented training set. The first step of the algorithm is executed towards initializing as an empty vector to store preprocessed images. In the next step, the algorithm processes each image x_i in the raw dataset D by applying resizing and standard preprocessing transformations to ensure consistency and uniformity. This initial data processing includes simplified preprocessing operations to ensure consistency and uniformity within the dataset, addressing common preprocessing challenges such as resizing images to standard dimensions and splitting the dataset into separate training, validation, and test sets. This automation enables the system to quickly and efficiently prepare data for model training. In addition, data augmentation techniques are applied to the training set to synthesize the diverse scenarios the model might encounter in real-time.

2.2. Deep predictive model

The proposed study considers the detection and classification of cervical cancer from Pap smear images as predictive problem. Let X represent the space of input images, where each image $x_i \in X$ is a Pap smear image. The goal is to map each x_i to a label $y_i \in Y$, where $Y=\{0,1\}$ denotes the binary classification space for non-cancerous (0) and cancerous (1) classes. In the case of multi-class classification, Y will be then including additional classes corresponding to different cervical cellular abnormalities. The predictive function $f:X\rightarrow Y$ is a DL model that takes an image x_i as input and predicts the corresponding label y^i. This function is parameterized by weights θ learned during the training process. Each DL model performs feature extraction by applying a series of transformations, such that for a given model m parameterized by θ_m, the feature extraction process can be represented as (1).

$$h_i^{(m)}=\phi^{(m)}(x_i;\theta m) \quad (1)$$

Where $h_i^{(m)}$ is the feature representation of the image x_i extracted by model m, and $\phi^{(m)}$ is the function defining the architecture of the model, which includes convolutional layers, activation functions, pooling layers, and fully connected layers. After feature extraction, the feature representation $hi^{(m)}$ is fed into a classifier $g^{(m)}$ to obtain the prediction $y^{i(m)}$ for the model m such that as in (2).

$$y^{i(m)}=g^{(m)}(h_i^{(m)};\psi_m) \quad (2)$$

Where ψ_m denotes the parameters of the classifier within model m. The proposed study considers training three different predictive models each with their specialized convolutive architecture and custom layers. The goal here is to harness the distinctive capabilities of these models to extract and learn from the complex patterns present in the data. The adopted predictive models are briefly described as follows:

- DenseNet201: the DenseNet201 model is a kind of advanced CNN which comes with a concept of dense connectivity where each layer connects directly to all subsequent layers in a feed-forward manner to capture better information flow and feature propagation throughout the network. This model basically introduced to address the vanishing gradient problem often encountered in conventional CNN models. The introduction of dense layers help maintains gradients and improve training in deeper networks. The DenseNet201 is trained on massive datasets of labeled images with 201 total layers, including convolutional layers, pooling layers, activation functions, and fully connected layers at the output layer.
- Xception: Xception CNN is another model for image classification task. This model utilizes depth wise separable convolutions to efficiently learn cross-channel correlations and spatial correlations in channels separately. Similar to ResNet models, this CNN model incorporates residual connections which allows the network to learn the difference between the input and the output of a block, facilitating the training process in deep models. In this proposed context this model allows it to act on complex input data efficiently, parsing through the intricacies of Pap smear images to identify and classify potential cancer indicators with a reduced computational cost compared to conventional convolutional approaches.
- Classical CNN: apart from implementing pre-trained model, the proposed study also implements a classical CNN model to train on the adopted dataset of Pap smear images. This model comprises a 5 2D convolutional layers, 5 batch normalization layers, 3 max pooling layers, 1 flatten layer, 3 dense layers, and 2 dropout layers for feature generalization extraction and classification purposes.

2.3. Proposed ensemble learning

Once each implemented DL model is trained, then they are incorporated into an ensemble learning framework where their predictive strengths are harnessed collectively. The ensemble method is distinct in that it incorporates an uncertainty-aware feature ranking and confidence aggregation mechanism. This mechanism critically evaluates each model's output, quantifying the confidence level of predictions and using this measure to dynamically influence the weight each model's prediction has on the final decision. The outcome is an aggregated predictive score that reflects a consensus among the models, weighted according to the certainty of their outputs.

The proposed ensemble learning scheme consider DL models, $M1$, $M2$, and $M3$ for a given set of test images $X=\{x_1, x_2, ..., x_n\}$, to predict their corresponding labels $Y=\{y_1, y_2, ..., y_n\}$. For each model $Mj=(1,2,3)$, predictions on X are obtained, resulting in confidence scores $Pj=\{pj1, pj2, ..., pjn\}$, where $pj1$ represents the probability distribution over possible classes for image xi predicted by model Mj. Afterwards, the proposed scheme computes confidence scores, such that: $Cj=\{cj1, cj2, ...,cjn\}$, from Pj, indicating the confidence of model Mj in its predictions for each xi. In this context, confidence is the probability assigned to the predicted class. The next operation involves computing an uncertainty measure Uji for each prediction pji, as an inverse operation related to its confidence score cji. Two distinct functions, $f1$ and $f2$, transform cji into ranks $Rj1i$ and $Rj2i$, respectively, which indicates the model's certainty in its prediction. The composite rank for each prediction is then determined by $Rji=f(Rj1i, Rj2i)$, where f is a function that combines the two ranks. The ensemble prediction for each image xi is obtained by aggregating the predictions from all models, weighted by their composite ranks Rji. The aggregated prediction for xi is given by $Paggi=g(R1i\cdot p1i, R2i\cdot p2i, R3i\cdot p3i)$, where g is an aggregation function that combines the weighted predictions, such as a weighted average or a weighted voting mechanism. The final class prediction, yi, is then determined based on $Paggi$. By adopting this approach, the proposed system ensures that the most reliable predictions are prioritized, thereby reducing the potential impact of any individual model's predictive anomalies. The result is a more accurate and dependable system for cervical cancer detection, one that can be seamlessly integrated into clinical workflows to assist medical professionals in diagnosing this condition more effectively.

3. RESULTS AND DISCUSSION

The design and development of the proposed system is carried out using python programming language executes in Anaconda distribution on Window 11 64-bit system. The training of the proposed DL model is carried out for 100 epochs with batch size 32 considering loss function categorical cross entropy and optimizer Adam. To achieve faster training process the model also uses GPU acceleration. The effectiveness of the proposed model is validated on the two dataset Sipakmed WSI [27] and Sipakmed cropped [28]. The performance of the model is assessed considering widely used classification metrics such as accuracy, precision, recall rate, and F1-score.

Figure 3 presents a visual analysis of the implemented predictive model on Sipakmed cropped considering training performance and testing performance. Based on the model training and validation graph it can be seen that the pre-trained models DenseNet and Xception both have smooth training and validation curve as compared to classical CNN. However classical CNN model slightly outperforms both DenseNet and Xception. This analysis can be also evident in confusion matrix where it can be clearly seen that the classical CNN has lower mis-classification errors along with high degree of precision.

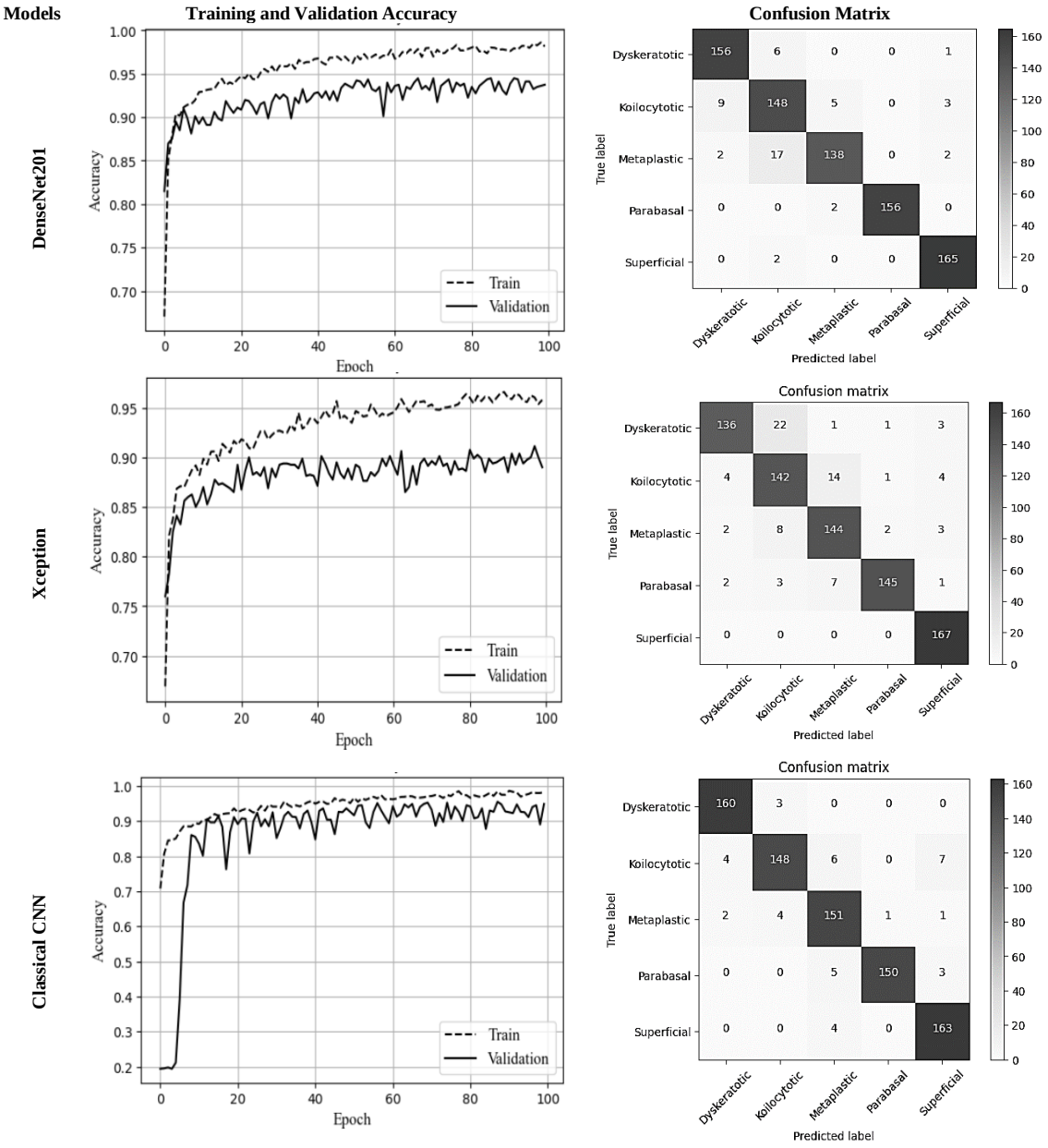


Figure 3. Visual analysis of the predictive model on Sipakmed cropped dataset

The Figure 4 presents a visual analysis of the implemented predictive model on Sipakmed WSI dataset considering training and testing performance. Similar trend can be also observed from validation curve that shows classical CNN agains has higher fluctuation compared to DenseNet201 and Xception model. However, both DenseNet201 and classical CNN model both shows higher training and validation accuracy. From the observation of confusion matrix, it can be concluded that while DenseNet201 and classical CNN show promising results with higher overall accuracy and more reliable class separation, the Xception model require further optimization to achieve comparable performance.

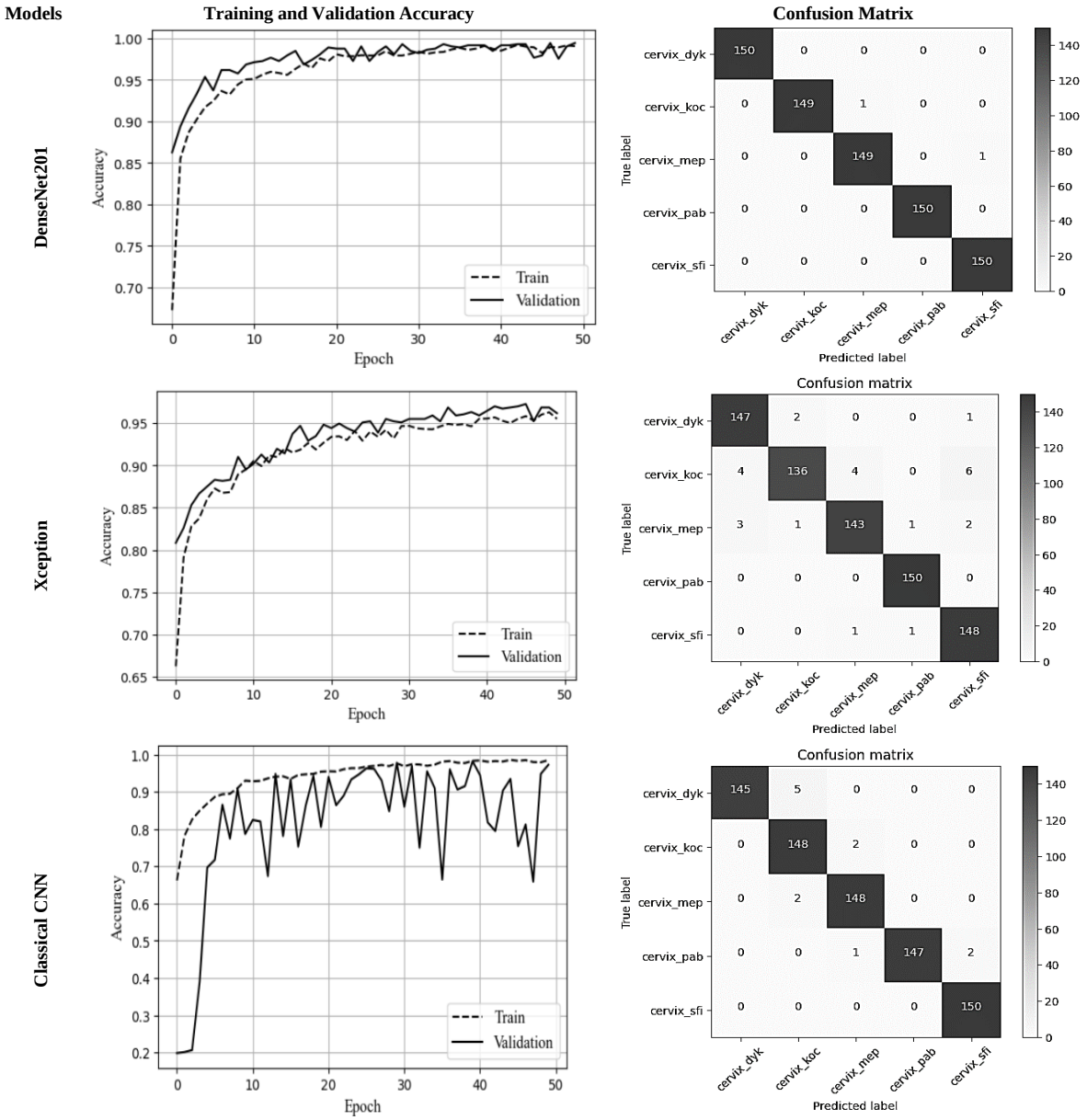


Figure 4. Visual analysis of the predictive model on Sipakmed WSI dataset

The numerical outcome presented in Table 2 shows comparison between various performance metrics for all three implemented DL models such as DenseNet201, Xception, a Classical CNN, and proposed ensemble method. The performance evaluation is done for Sipakmed cropped dataset of cervical cancer cell classification task. The evaluation metrics include precision, recall, and F1-score for each of the five cell types such as dyskeratotic, koilocytotic, metaplastic, parabasal, and superficial. It can be seen that the individual models offer competitive performance but with slight variations, whereas the proposed ensemble method outperforms each individual models. The outcome validates the effectiveness of our

ensemble model with an enhanced prediction accuracy with overall highest accuracy for most cervical categories.

Table 2. Analysis of numerical outcome on Sipakmed cropped dataset

Models	Metrics	Dyskeratotic	Koilocytotic	Metaplastic	Parabasal	Superficial
DenseNet201	Precision	0.93	0.86	0.95	1.00	0.96
	Recall	0.96	0.90	0.87	0.99	0.99
	F1	0.95	0.88	0.91	0.99	0.98
	Overall accuracy			94		
Xception	Precision	0.94	0.81	0.87	0.97	0.94
	Recall	0.83	0.86	0.91	0.92	1.00
	F1	0.89	0.84	0.89	0.94	0.97
	Overall accuracy			0.90		
Classical CNN	Precision	0.96	0.95	0.91	0.99	0.94
	Recall	0.97	0.90	0.95	0.95	0.98
	F1	0.98	0.92	0.93	0.97	0.96
	Overall accuracy			0.95		
Proposed ensemble	Precision	0.97	0.96	0.94	1.00	0.97
	Recall	0.97	0.90	0.97	0.99	1.00
	F1	0.97	0.93	0.96	0.99	0.99
	Overall accuracy			0.97		

The quantitative analysis in Table 3 shows the comparison between the implemented prediction models considering various performance metrics evaluated on the Sipakmed WSI dataset. It can be seen that the classic CNN model achieved similar performance to DenseNet201 in terms of accuracy, with full scores obtained for all cell types except all cells. For dyskeratosis and parabasal cells, the recall was slightly lower but still high. With an overall accuracy of 98%, the classic CNN proved to be very effective and comparable to more complex models. It can also be seen that the proposed ensemble method shows excellent performance, matching DenseNet201 and classic CNN models with perfect precision and recall for almost all cell types. The F1-score confirms the model's balanced precision and recall, resulting in perfect classification results. The overall accuracy was a perfect 100%, making the ensemble method the most reliable and robust method for testing models.

Table 3. Analysis of numerical outcome on Sipakmed WSI dataset

Models	Metrics	Dyskeratotic	Koilocytotic	Metaplastic	Parabasal	Superficial
DenseNet201	Precision	1.00	1.00	0.99	1.00	0.99
	Recall	1.00	0.99	0.99	1.00	1.00
	F1	1.00	1.00	0.99	1.00	1.00
	Overall accuracy			1.00		
Xception	Precision	0.95	0.98	0.97	0.99	0.94
	Recall	0.98	0.91	0.95	1.00	0.99
	F1	0.97	0.94	0.96	0.99	0.96
	Overall accuracy			0.97		
Classical CNN	Precision	1.00	0.95	0.98	1.00	0.99
	Recall	0.97	0.99	0.99	0.98	1.00
	F1	0.98	0.97	0.98	0.99	0.99
	Overall accuracy			0.98		
Proposed Ensemble	Precision	1.00	1.00	1.00	1.00	0.99
	Recall	1.00	0.99	1.00	1.00	1.00
	F1	1.00	1.00	1.00	1.00	1.00
	Overall accuracy			1.00		

4. CONCLUSION

This paper introduces the design and development of a predictive function for the detection and classification of cervical cancer from Pap smear images. The main contribution of this work is the development of an advanced ensemble learning technique that incorporates uncertainty-aware prediction ranking and confidence aggregation to maximize the consensus reliability of different image classification models, thereby reducing the impact of uncertain predictions. The experimental result demonstrates that the proposed ensemble learning method driven by the uncertainty-aware decision mechanism achieved better diagnostic performance than traditional single-model approaches. Our findings show that the proposed scheme overcomes the limitation of using a single model and provides a promising approach to clinicians in

early detection of cervical cancer. The proposed uncertainty-aware prediction ranking and confidence aggregation strategy is our main contribution which is also adaptive and flexible in its design to integrate more different model predictions in a coherent manner to provide reliable results. In the future, the proposed study can be extended towards adopting dynamic and adaptive data augmentation technique which will be selectively applied based on the dataset preprocessing requirement. In addition, the proposed study also extends this model towards detecting more sophisticated disease such as like Alzheimer's and autism spectrum disorder following effective feature engineering techniques.

ACKNOWLEDGMENTS

We are very thankful towards the dedicated effort put forward by the staff members deputed in Research Center, Sai Vidya Institute of Technology and Vision and Learning Laboratory of Presidency University. Their assistance towards facilitating data, computing environment for cluster image processing proved very helpful for this research work.

REFERENCES

[1] A. Kumar *et al.*, "Nanotherapeutics: a novel and powerful approach in modern healthcare system," in *Nanotechnology in Modern Animal Biotechnology*, Elsevier, 2019, pp. 149–161, doi: 10.1016/B978-0-12-818823-1.00010-7.
[2] P. A. Cohen, A. Jhingran, A. Oaknin, and L. Denny, "Cervical cancer," *The Lancet*, vol. 393, no. 10167, pp. 169–182, 2019, doi: 10.1016/S0140-6736(18)32470-X.
[3] L. Peirson, D. Fitzpatrick-Lewis, D. Ciliska, and R. Warren, "Screening for cervical cancer: a systematic review and meta-analysis," *Systematic reviews*, vol. 2, pp. 1–14, 2013, doi: 10.1186/2046-4053-2-35.
[4] A. D. Juan *et al.*, "SEOM clinical guidelines for cervical cancer (2019)," *Clinical and Translational Oncology*, vol. 22, no. 2, pp. 270–278, 2020, doi: 10.1007/s12094-019-02271-z.
[5] S. Finocchario-Kessler, C. Wexler, M. Maloba, N. Mabachi, F. Ndikum-Moffor, and E. Bukusi, "Cervical cancer prevention and treatment research in Africa: a systematic review from a public health perspective," *BMC Women's Health*, vol. 16, no. 1, 2016, doi: 10.1186/s12905-016-0306-6.
[6] S. L. Bedell, L. S. Goldstein, A. R. Goldstein, and A. T. Goldstein, "Cervical cancer screening: past, present, and future," *Sexual Medicine Reviews*, vol. 8, no. 1, pp. 28–37, 2020, doi: 10.1016/j.sxmr.2019.09.005.
[7] N. Wentzensen and M. V. K. Doeberitz, "Biomarkers in cervical cancer screening," *Disease Markers*, vol. 23, no. 4, pp. 315–330, 2007, doi: 10.1155/2007/678793.
[8] A. Jeffries *et al.*, "Cervical visual inspection with acetic acid (VIA) and oncogenic human papillomavirus screening in rural indigenous guatemalan women: time to rethink VIA," *International Journal of Environmental Research and Public Health*, vol. 18, no. 23, 2021, doi: 10.3390/ijerph182312406.
[9] W. William, A. H. Basaza-Ejiri, J. Obungoloch, and A. Ware, "A review of applications of image analysis and machine learning techniques in automated diagnosis and classification of cervical cancer from pap-smear images," *2018 IST-Africa Week Conference, IST-Africa 2018*, vol. 164, pp. 15–22, 2018.
[10] K. B. Johnson *et al.*, "Precision medicine, AI, and the future of personalized health care," *Clinical and Translational Science*, vol. 14, no. 1, pp. 86–93, 2021, doi: 10.1111/cts.12884.
[11] D. Painuli, S. Bhardwaj, and U. Köse, "Recent advancement in cancer diagnosis using machine learning and deep learning techniques: A comprehensive review," *Computers in Biology and Medicine*, vol. 146, 2022, doi: 10.1016/j.compbiomed.2022.105580.
[12] P. Xue *et al.*, "Deep learning in image-based breast and cervical cancer detection: a systematic review and meta-analysis," *NPJ Digital Medicine*, vol. 5, no. 1, 2022, doi: 10.1038/s41746-022-00559-z.
[13] V. S. Palanisamy, R. K. Athiappan, and T. Nagalingam, "Pap smear based cervical cancer detection using residual neural networks deep learning architecture," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 4, 2022, doi: 10.1002/cpe.6608.
[14] M. Kalbhor, S. Shinde, H. Joshi, and P. Wajire, "Pap smear-based cervical cancer detection using hybrid deep learning and performance evaluation," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization*, vol. 11, no. 5, pp. 1615–1624, 2023, doi: 10.1080/21681163.2022.2163704.
[15] 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*, vol. 15, pp. 163–176, 2022, doi: 10.2147/MDER.S366303.
[16] S. Yu *et al.*, "Automatic classification of cervical cells using deep learning method," *IEEE Access*, vol. 9, pp. 32559–32568, 2021, doi: 10.1109/ACCESS.2021.3060447.
[17] A. Ghoneim, G. Muhammad, and M. S. Hossain, "Cervical cancer classification using convolutional neural networks and extreme learning machines," *Future Generation Computer Systems*, vol. 102, pp. 643–649, 2020, doi: 10.1016/j.future.2019.09.015.
[18] S. Cheng *et al.*, "Robust whole slide image analysis for cervical cancer screening using deep learning," *Nature Communications*, vol. 12, no. 1, 2021, doi: 10.1038/s41467-021-25296-x.
[19] S. Alsubai *et al.*, "Privacy preserved cervical cancer detection using convolutional neural networks applied to Pap smear images," *Computational and Mathematical Methods in Medicine*, vol. 2023, no. 1, 2023, doi: 10.1155/2023/9676206.
[20] N. K. Chauhan, K. Singh, A. Kumar, and S. B. Kolambakar, "HDFCN: A robust hybrid deep network based on feature concatenation for cervical cancer diagnosis on WSI Pap smear slides," *BioMed Research International*, vol. 2023, no. 1, 2023, doi: 10.1155/2023/4214817.
[21] M. Alsalatie, H. Alquran, W. A. Mustafa, Y. M. Yacob, and A. A. Alayed, "Analysis of cytology Pap smear images based on ensemble deep learning approach," *Diagnostics*, vol. 12, no. 11, 2022, doi: 10.3390/diagnostics12112756.
[22] Q. M. Ilyas and M. Ahmad, "An enhanced ensemble diagnosis of cervical cancer: a pursuit of machine intelligence towards sustainable health," *IEEE Access*, vol. 9, pp. 12374–12388, 2021, doi: 10.1109/ACCESS.2021.3049165.
[23] K. Chadaga, S. Prabhu, N. Sampathila, R. Chadaga, S. K. S., and S. Sengupta, "Predicting cervical cancer biopsy results using demographic and epidemiological parameters: a custom stacked ensemble machine learning approach," *Cogent Engineering*, vol. 9, no. 1, 2022, doi: 10.1080/23311916.2022.2143040.

[24] S. Gupta and M. K. Gupta, "Computational prediction of cervical cancer diagnosis using ensemble-based classification algorithm," *The Computer Journal*, vol. 65, no. 6, pp. 1527–1539, 2022, doi: 10.1093/comjnl/bxaa198.
[25] B. S. Deo, M. Pal, P. K. Panigrahi, and A. Pradhan, "CerviFormer: a Pap smear-based cervical cancer classification method using cross-attention and latent transformer," *International Journal of Imaging Systems and Technology*, vol. 34, no. 2, 2024, doi: 10.1002/ima.23043.
[26] P. Mahajan and P. Kaur, "Improving cervical cancer classification in Pap smear images with enhanced segmentation and deep progressive learning-based techniques," *Diagnostic Cytopathology*, vol. 52, no. 6, pp. 313–324, 2024, doi: 10.1002/dc.25295.
[27] U. J. Patel, "SIPaKMeD WSI dataset," *Kaggle*. 2020. Accessed: Mar. 07, 2024. [Online]. Available: https://www.kaggle.com/datasets/umangjpatel/sipakmed-wsi
[28] M. E. Plissiti, P. Dimitrakopoulos, G. Sfikas, C. Nikou, O. Krikoni, and A. Charchanti, "Sipakmed: A new dataset for feature and image based classification of normal and pathological cervical cells in Pap smear images," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, 2018, pp. 3144–3148, doi: 10.1109/ICIP.2018.8451588.

BIOGRAPHIES OF AUTHORS

Sreelatha holds the bachelors and master's degree from Visvesvaraya Technological University. She is currently working as an assistant professor in Department of Computer Science and Engineering at Presidency University, Bangalore. Her research area includes image processing and machine learning. She has published over 5 papers in international journals and conferences. She can be contacted at email: sreelatha.sajeev2@gmail.com or sreelatha.pk@presidencyuniversity.in.

Dr. Vrinda Shivashetty received the doctorate degree in computer science and engineering from Gulbarga University. She is currently working as a Professor in Department of Information Science and Engineering at Sai Vidya Institute of Technology, Bangalore. She is recipient of many awards like "emerging leader in engineering", powered by Venus International Foundation in 2018 at Chennai, "outstanding women in engineering", powered by Venus International Foundation in 2018 at Chennai, "senior woman educator and scholar award", powered by NFED Coimbatore in 2017, and also "excellent teacher" award 5 times for achieving 100% results in university exams. She can be contacted at email: hodise@saividya.ac.in.