An Explainable Deep Learning Framework for Cervical Squamous Cell Classification using Transfer Learning, Ensemble Learning, and Knowledge Distillation via Teacher Assistant

Umar Hasan[1][0009−0009−6659−2006], Abdullah Al Raiyan[2], and Sifat Haque Zidan[3]

North South University, Dhaka 1229, Bangladesh
[1]umar.hasan@northsouth.edu
[2]abdullah.raiyan@northsouth.edu
[3]sifat.zidan@northsouth.edu

Abstract. Cervical cancer remains a significant global health challenge, particularly in resource-limited settings. Microscopic examination of Pap smear slides is the standard screening method, but is labor-intensive and requires expert cytopathologists. Computer-Aided Diagnosis systems leveraging deep learning offer a promising solution for automating and improving the accuracy of cervical cell classification. This paper presents an explainable deep learning framework designed to classify cervical squamous cells from the SipakMed dataset into five distinct classes representing cancer severity. Our approach integrates several advanced techniques: (1) Transfer learning, fine-tuning multiple state-of-the-art pretrained convolutional neural networks using differential learning rates; (2) Ensemble learning, combining the predictions of the fine-tuned models using weighted averaging to enhance robustness and accuracy; (3) Knowledge Distillation via a Teacher Assistant (TA), where knowledge from a large ensemble teacher model is distilled into a medium-sized TA model (SqueezeNet), which subsequently teaches a smaller, more efficient student model (AgileNet, our custom model); and (4) Explainable AI, employing Grad-CAM++ to provide visual explanations for model predictions, enhancing transparency and trustworthiness. Our proposed ensemble model achieves a test accuracy of 98.01%. Furthermore, the distilled Teacher Assistant model (SqueezeNet) and the distilled Student model (AgileNet) achieve a commendable accuracy of 97.21% and 97.61%, respectively, demonstrating the effectiveness of the TA-KD framework in creating efficient yet accurate models. The integration of explainable AI methods provides insights into the Student models' decision-making processes. This framework represents a substantial advancement in automated cervical squamous cell classification, offering improved performance and interpretability.

Keywords: Cervical Cancer · Deep Learning · Transfer Learning · Ensemble Learning · Knowledge Distillation · Teacher Assistant · Explainable AI · Grad-CAM++ · Image Classification · Pap Smear · SipakMed · AgileNet · SqueezeNet.

1 Introduction

Cervical cancer, originating in the cells of the cervix, poses a significant threat to women's health globally. It represents a major cause of morbidity and mortality among women worldwide, with a disproportionate impact in developing nations [2]. In Bangladesh, cervical cancer ranks as the second leading malignancy among females, accounting for 12% of cases [7]. Statistics from 2018 indicated 8,068 new diagnoses (an incidence rate of 10.6 per 100,000 women) and 5,214 fatalities (a mortality rate of 7.1 per 100,000 women) [7]. Projections suggest that without intervention, cumulative deaths from cervical cancer in Bangladesh could reach 505,703 by 2070, potentially increasing to 1,042,859 by 2120 [9]. Microscopic examination of Pap smears is the standard for detecting potentially cancerous changes in the cervix [1]. However, the manual interpretation of Pap smears is inherently time-consuming and subjective, demanding highly trained cytopathologists. This poses significant challenges for large-scale screening initiatives, particularly in regions where expert resources are scarce, potentially leading to delays in diagnosis and treatment [2]. This highlights the need for efficient Computer-Aided Diagnosis (CAD) systems. Early detection through screening programs, predominantly using the Papanicolaou (Pap) smear test, is paramount for effective treatment and improving patient outcomes [2].

Deep learning, particularly Convolutional Neural Networks (CNNs), has shown great potential in medical image analysis, including cervical cancer screening. These models can automatically learn features from images for accurate classification. Existing work has demonstrated the potential of CNNs, often combined with transfer learning from models pre-trained on large datasets like ImageNet, to automatically extract features and classify cervical cell images [**?**, **?**]. However, these approaches face persistent challenges. Capturing the subtle morphological variations distinguishing different cell classes remains difficult, often requiring very deep or complex models. Publicly available datasets for training, like SipakMed [32], while valuable, are often limited in size compared to general computer vision datasets, increasing the risk of overfitting. Furthermore, many existing deep learning solutions suffer from a lack of transparency, operating as "black boxes" that hinder clinical acceptance. Ensemble methods, such as the one proposed by Gangrade et al. [2] achieving 94% accuracy by combining CNN, AlexNet, and SqueezeNet, aim to improve robustness and accuracy, yet there is still a clear need for frameworks that push performance boundaries further while simultaneously addressing model efficiency and interoperability.

This paper proposes an advanced, explainable deep learning framework specifically designed to overcome these limitations in cervical squamous cell classification. Our approach aims to deliver enhanced accuracy, improved computational efficiency through model compression, and increased transparency via explainability techniques. The primary contributions of this work are:

- **Enhanced Transfer Learning:** Fine-tuning a diverse set of state-of-the-art pre-trained CNN architectures using differential learning rates for optimal feature extraction from the SipakMed dataset.

- **Improved Ensemble Learning:** Developing a weighted averaging ensemble of the fine-tuned models, demonstrating superior classification accuracy compared to individual models and the baseline ensemble in [2].
- **Knowledge Distillation via Teacher Assistant (TA-KD):** Implementing a novel Knowledge Distillation (KD) strategy where a complex ensemble acts as the teacher, distilling knowledge into a compact SqueezeNet Teacher Assistant (TA) model, which then trains an even smaller, custom student model (AgileNet) for efficient deployment.
- **Explainable AI (XAI) Integration:** Employing the Grad-CAM++ technique to visualize the regions within Pap smear images that the student model (AgileNet) focuses on, providing interpretability and increasing trust in the diagnostic predictions.

Our framework achieves high performance on the SipakMed dataset and provides insights into the model's decision process through XAI, representing a significant step towards more accurate, efficient, and trustworthy CAD systems for cervical cancer screening. The paper is structured as follows:

- Section II provides an overview of the literature related to cervical cancer classification, deep learning approaches, Knowledge Distillation, and Explainable AI in medical imaging.
- Section III outlines the proposed methodology, detailing the dataset, the architectures of the teacher (Weighted Ensemble), teacher assistant (SqueezeNet), and student (AgileNet - custom CNN) models, the Knowledge Distillation process, the GradCAM++ implementation, description of baseline models, and ensemble configurations used for comparison. It also presents the experimental setup and evaluation metrics.
- Section IV focuses on the quantitative results, including performance comparisons between the student model (with/without KD), teacher assistant model, teacher ensemble, baseline architectures, the full ensemble model, and the reference study [2]. It also presents the Explainability Analysis, interpreting the GradCAM++ visualizations for the student model.
- Section V discusses the overall findings, highlighting the effectiveness of the proposed TA-KD approach, the significance of outperforming the reference study, insights from the XAI analysis, limitations, and potential future research directions.
- Section VI offers concluding remarks, summarizing the key contributions and implications of the study.

2 Related Works

This literature review examines a wide array of techniques and models applied to classifying and detecting cervical cancer. It covers diverse approaches, including computer algorithms, deep learning strategies, and ensemble methods, highlighting the variety of methodologies, datasets, and models used across different studies to obtain accurate outcomes.

2.1 Early Computational Methods and Deep Learning Foundations

Early efforts explored various computational approaches for cervical cancer detection. Mango et al. combined traditional Pap tests with artificial neural networks (ANN) [12], while Sukumar and Gnanamurthy utilized MRI scans with hybrid SVM and neuro-fuzzy classifiers [13]. The advent of deep learning brought significant advancements. Bora et al. applied deep CNNs for image identification, improving accuracy via feature selection [14], and Hyeon et al. demonstrated the effectiveness of CNNs like VGG16 for feature extraction from cervical MRIs [15]. Sanyal et al. also developed a CNN for finding abnormal areas in conventional smears with high diagnostic accuracy [17]. Complementary techniques like surface-enhanced Raman scattering (SERS) combined with machine learning also showed promise for predicting sample pathology [18]. These foundational works established the potential of computational methods, particularly deep learning, in analyzing complex medical data like Pap smears and MRIs. These studies highlight the power of CNNs for feature extraction and classification in this domain, inspiring our use of established CNN architectures. While promising, these early methods often required significant feature engineering or faced limitations in handling the complexity and variability of cytological images, indicating a need for more robust and automated feature learning, such as that offered by advanced transfer learning.

2.2 Advanced Deep Learning, Transfer Learning, and Ensemble Techniques

Subsequent research focused on leveraging more sophisticated deep learning models and techniques. Promworn et al. compared several deep learning models on cytopathology images, identifying DenseNet161 as a top performer, which also inspired the efficient ColpoNet [16,31]. The utility of transfer learning, using pre-trained CNNs like AlexNet and VGG-16, became evident, with studies like Taha et al. [19] and Kudva et al. [20] showing notable accuracy improvements by leveraging pre-trained features. Ensemble methods further boosted performance; Xue et al. used Ensemble Transfer Learning (ETL) for histopathology images [21], while Chen et al. also achieved high accuracy with CNNs and transfer learning [22]. Ghoneim et al. combined CNNs with ELM classifiers [23], and Arifianto et al. applied CNNs to diverse datasets [24], both yielding significant results. Hussain et al. proposed several DCNN models with high accuracy [25]. Kang et al. combined Raman spectroscopy with a novel hierarchical CNN (H-CNN), surpassing traditional methods for classifying cancer stages [26]. Pacal et al. achieved state-of-the-art results using Vision Transformers (ViT) and CNNs with data augmentation and ensembles [28], while Pramanik et al. introduced a fuzzy distance-based ensemble method [29]. Gangrade et al. specifically used an ensemble of CNN, AlexNet, and SqueezeNet on the SipakMed dataset, achieving 94% accuracy [2] (cited as [2] in the user template, using [2] for consistency). Integrating multiple models through ensemble learning, as noted in [30], consistently shows potential for higher accuracy and robustness compared to single models. The success of transfer learning with diverse architectures (DenseNet, AlexNet, VGG, ViT) and the proven benefit of ensemble methods, including the work by Gangrade et al. [2], directly inspired our approach of fine-tuning multiple SOTA models and combining them using a weighted ensemble. While ensemble methods improve accuracy, they often result in large, computationally expensive models. Furthermore, even with advanced models, achieving near-perfect accuracy and ensuring generalization remains a challenge, as highlighted by reviews like Youneszade et al. [27]. There is a gap in developing methods that maintain or exceed state-of-the-art ensemble accuracy while simultaneously creating more computationally efficient models suitable for wider deployment.

2.3 Knowledge Distillation and Explainable AI

To address the challenge of large model sizes, Knowledge Distillation (KD), introduced by Hinton et al. [36], offers a compelling compression technique. It trains a smaller 'student' network to mimic the outputs ('soft targets') of a larger 'teacher' network, transferring knowledge beyond ground truth labels. KD has been successfully applied in various fields, including medical imaging, enabling efficient yet powerful models. Concurrently, the need for transparency in medical AI necessitates Explainable AI (XAI). Techniques like Grad-CAM [34] and its successor GradCAM++ [33] provide visual explanations by generating heatmaps highlighting influential input regions for a model's decision. This is crucial for building trust and understanding model behavior in critical healthcare applications. The principles of KD provide a clear pathway to model compression, inspiring our use of the TA-KD framework. The necessity for transparency in medical diagnosis, addressed by XAI techniques like GradCAM++, motivated its integration into our workflow. While KD and XAI are established fields, their synergistic application specifically within a Teacher-Assistant framework for cervical cell classification, aimed at creating both highly accurate, efficient, *and* interpretable models, represents an underexplored area. Evaluating how distillation affects the explanations generated by XAI methods is also an important consideration.

The present work strategically builds upon these established research areas. We apply Knowledge Distillation through a Teacher-Assistant framework to train an efficient custom student CNN architecture (AgileNet) tailored for cervical cell classification, using a SqueezeNet TA trained by a top-performing ensemble teacher. A key contribution involves rigorously comparing our distilled models against individual fine-tuned models, the ensemble teacher, and the reference study [2]. Additionally, we leverage GradCAM++ [33] to provide crucial explainability for our student model's predictions, analyzing its decision-making process.

For a consolidated overview of the key studies discussed, Table 1 summarizes the relevant literature, categorized by the proposed method, the dataset utilized in the study, and the primary results attained.

Table 1: Summary of Cervical Cancer Detection Literature

Reference	Method	Dataset	Results
Minge et al. [12]	Pap smear test + ANN model	N/A	N/A
Sukumar & Gnanamurthy [13]	MRI scans + SVM + NN	Herlev data	99.1% acc. (2-class)
Bora et al. [14]	CNN-based classification, feature selection	Private dataset	Improved accuracy
Hyeon et al. [15]	CNNs + VGG16 feature extraction	7134 MRIs	SVM F1 score superior
Promworn et al. [31]	Comparative analysis of models	N/A	DenseNet161: 94.38% acc.
ColpoNet [16]	Inspired by DenseNet	Nat. Cancer Institute dataset	81.35% acc.
Parikshit Sanyal et al. [17]	CNN for detecting abnormal foci	1838 microphotographs	95.46% diagnosis acc.
Karunakaran et al. [18]	Ultrasensitive SERS	Cervix cell samples	Avg. acc. 95.46%
Taha et al. [19]	Pre-trained CNN architecture	Herlev dataset	99.19% acc. (2-class)
Kudva et al. [20]	Hybrid transfer learning (AlexNet, VGG-16)	N/A	91.46% classification acc.
Xue et al. [21]	Ensemble Transfer Learning (ETL)	Herlev dataset	Highest acc. 98.61%
Chen et al. [22]	Fine-tuned CNN architectures	4993 histology images	97.42% classification acc.
Ghoneim et al. [23]	CNNs + ELM classifiers	Herlev database	99.5% detection acc., 91.2% classification acc.
Kang et al. [26]	Raman spectroscopy + H-CNN	Tissue samples	>94% acc. classifying tissues
Youneszade et al. [27]	Review	Review	Overview of DL techniques
Pacal et al. [28]	ViT, CNN-based models, ensemble	Massive dataset	Record-breaking accuracy
Pramanik et al. [29]	Fuzzy distance-based ensemble	Pap smear images	Promising results
Gangrade et al. [2]	Ensemble (CNN, AlexNet, SqueezeNet)	SipakMed Dataset	94% acc. (5-class)

3 Methodology

This section details the proposed framework, outlining the data handling, core techniques, and model architectures employed.

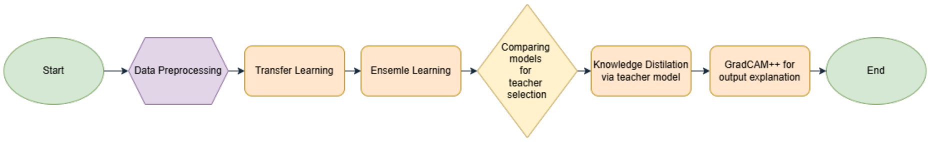

Start
Data Preprocessing
Transfer Learning
Ensemle Learning
Comparing models for teacher selection
Knowledge Distilation via teacher model
GradCAM++ for output explanation
End

Fig. 1: Research Experiment Methodology

3.1 Data and Preprocessing

We utilized the SipakMed dataset [32], publicly available on Kaggle [**?**]. This dataset contains 5015 digitized conventional Pap smear images, manually cropped by experts. The images are categorized into five classes based on squamous cell type, representing different stages of cervical cancer severity: Dyskeratotic (im_Dyskeratotic), Koilocytotic (im_Koilocytotic), Metaplastic (im_Metaplastic), Parabasal (im_Parabasal), and Superficial-Intermediate (im_Superficial-Intermediate). Following the preprocessing steps similar to those in [2] and adapted in our notebook, all images were resized to a uniform size of 64x64 pixels.

- **Training Set Augmentation:** To improve model generalization and reduce overfitting, random horizontal flipping (p=0.5) was applied to the training images.
- **Normalization:** All images (training, validation, and test) were converted to PyTorch tensors and normalized using the ImageNet mean ([0.485, 0.456, 0.406]) and standard deviation ([0.229, 0.224, 0.225]), as required by the pre-trained models.
- **Data Splitting:** The dataset was split into training, validation, and testing sets using a fixed random seed (RANDOM_SEED = 42) for reproducibility. The splits were allocated as follows: 70% for training (3510 images), 15% for validation (752 images), and 15% for testing (753 images). The validation set was used for hyperparameter tuning and selecting the best model during training, while the test set was held out for final model evaluation.

3.2 Ensemble Learning Pipeline

To establish a high-performance benchmark and provide a comparison point, a weighted ensemble model was created by combining the predictions of the highest-performing fine-tuned individual models. The pipeline involved:

1. **Model Selection:** Selecting the 11 best-performing individual models fine-tuned via transfer learning: AlexNet, ResNet-50, ResNet-101, ResNet-152, DenseNet-169, DenseNet-201, VGG-16, VGG-19, EfficientNet-B3, ConvNeXt-Tiny, and RegNetY-16GF. (SqueezeNet was excluded due to lower baseline performance).
2. **Weight Calculation:** Determining weights for each model (M_i) based on their respective test accuracies ($test\_acc_i$) reported in Section 4. The weight w_i is calculated proportionally:

   $$w_i = \frac{test\_acc_i}{\sum_{j=1}^{11} test\_acc_j} \tag{1}$$

   where the sum is over the 11 selected models.
3. **Prediction Aggregation:** For a given test input image x, let $p_i(x)$ be the probability vector (output of the softmax layer) from model M_i. The final ensemble prediction $P_{ens}(x)$ is the weighted average of these probabilities:

   $$P_{ens}(x) = \sum_{i=1}^{11} w_i p_i(x) \tag{2}$$

   The final predicted class corresponds to the class with the highest probability in $P_{ens}(x)$. This 11-model ensemble serves as our highest-performing benchmark. Note that a different ensemble (Top-5 models) was used as the Teacher for KD, as described previously.

3.3 Revisiting Knowledge Distillation via Teacher Assistant (TA-KD Pipeline)

To create computationally efficient models while maintaining high accuracy, we implemented a knowledge distillation strategy using a Teacher Assistant (TA) [35], following this pipeline:

1. **Teacher Model Selection:** A powerful ensemble model (Teacher) was formed using the 11 high-performing fine-tuned pre-trained models identified from the initial transfer learning phase (AlexNet, Resnet50, ResNet101, ResNet152, DenseNet169, DenseNet201, VGG16, VGG19, EfficientNet B3, ConvNeXt Tiny, and RegNet Y 16GF). The ensemble's aggregated prediction serves as the source of rich 'dark knowledge'.
2. **TA Model Training:** A compact pre-trained model, SqueezeNet-v1.1 (fine-tuned), was chosen as the TA. The TA was trained to mimic the output probabilities (soft labels) of the Teacher ensemble using a KD loss function, combining Kullback-Leibler (KL) divergence with the standard cross-entropy loss:

   $$L_{KD\_TA} = \alpha L_{CE}(y, p_{TA}) + (1-\alpha) L_{KL}(p_T, p_{TA}) \tag{3}$$

   where p_T and p_{TA} are the soft predictions (using temperature scaling, T) of the Teacher and TA, y is the true label, L_{CE} is cross-entropy loss, L_{KL} is KL divergence loss, and α is a weighting factor balancing the two loss components.

3. **Student Model Training:** A smaller, custom lightweight model, AgileNet, was designed as the Student. The Student was then trained using the trained TA model (SqueezeNet-v1.1) as its teacher, applying a similar KD loss function:

$$L_{KD\_Student} = \beta L_{CE}(y, p_S) + (1 - \beta) L_{KL}(p'_{TA}, p_S) \tag{4}$$

where p'_{TA} and p_S are the soft predictions of the TA and Student (using temperature T), and β is a weighting factor.

This hierarchical distillation pipeline allows the highly compact Student model (AgileNet) to learn effectively from the distilled knowledge passed down from the large Teacher ensemble via the intermediate TA (SqueezeNet-v1.1).

3.4 Model Architectures

This section details the architectures involved in our framework.

Baseline and Transfer Learning Models: We employed transfer learning by utilizing several CNN architectures pre-trained on the ImageNet dataset: AlexNet, SqueezeNet-v1.1, ResNet (50, 101, 152), VGG (16, 19), DenseNet (169, 201), EfficientNet-B3, ConvNeXt-Tiny, and RegNet-Y-16GF [?]. The fine-tuning strategy involved replacing the final classifier, unfreezing final layers, and using differential learning rates with the AdamW optimizer [?] and Cross-Entropy Loss. These fine-tuned models serve as baselines for comparison.

Teacher Ensemble Architecture: The Teacher model used for KD is a weighted ensemble average of 11 high-performing fine-tuned models including AlexNet, ResNet-50, ResNet-101, ResNet-152, VGG-16, VGG-19, DenseNet-169, DenseNet-201, EfficientNet-B3, ConvNeXt-Tiny, and RegNet-Y-16GF. It does not have a distinct architecture itself but aggregates the outputs of its constituent models.

Teacher Assistant (TA) Architecture: SqueezeNet-v1.1 [?] was used as the TA. It is known for its computational efficiency, achieved through the use of 'fire modules' consisting of a squeeze convolution layer (1x1 filters) feeding into an expand layer (mix of 1x1 and 3x3 filters). For its role as TA, we used a version pre-trained on ImageNet and subsequently fine-tuned specifically for the SipakMed classification task using the differential learning rate strategy. This fine-tuning involved unfreezing the final convolutional classifier layer (replaced for 5 classes) and the last three Fire modules (features 10, 11, and 12) while keeping earlier layers frozen. This adapted model served as the starting point for the KD training of the TA.

Student Architecture (AgileNet): We designed AgileNet as a custom lightweight CNN architecture optimized for efficiency and performance on this specific task after distillation. It comprises 286,229 parameters. The architecture takes a 3x64x64 image as input and consists of three main convolutional blocks followed by fully connected layers. Each convolutional block uses a 3x3 convolutional layer (with padding 1 and no bias) followed by Batch Normalization [?] and a ReLU activation function [?], concluded by 2x2 Max Pooling for spatial downsampling. The channel dimensions increase through the blocks: Block 1 (3 channels in, 16 out), Block 2 (16 channels in, 32 out), Block 3 (32 channels in, 64 out). After the third pooling layer, the feature map (64x8x8) is flattened to a vector of size 4096. This vector feeds into a fully connected layer reducing the dimension to 64, followed by ReLU activation and a Dropout layer [?] with a probability of 0.3 for regularization. The final output layer is a fully connected layer mapping the 64 features to the number of output classes (5). [Placeholder: Add citations for BatchNorm [?], ReLU [?], Dropout [?]]

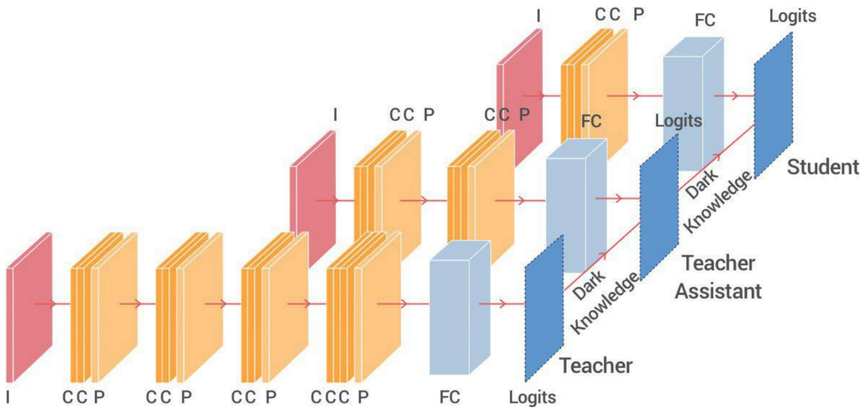


Fig. 2: TA fills the gap between student & teacher

Explainable AI (XAI) Implementation: To understand the decision-making process of our distilled student model (AgileNet), we integrated Grad-CAM++ [33]. We utilized the implementation from the ‘pytorch-grad-cam‘ library [34]. Heatmaps were generated by targeting the final convolutional layer of AgileNet (‘conv3‘). Visualizations were produced for representative, correctly classified samples selected from the test set for each of the five classes to illustrate the model’s focus areas leading to its predictions.

4 Experiments and Results

This section presents the experimental setup, evaluation methodology, results, and analysis of the proposed framework.

4.1 Setup

All experiments were conducted within a consistent environment to ensure reproducibility and fair comparison.

Implementation Details: All models were implemented using PyTorch version [placeholder: e.g., 1.13+] and trained on a Kaggle environment equipped with an NVIDIA Tesla T4 GPU. Key libraries included torchvision, timm, scikit-learn, and Grad-CAM [?]. The AdamW optimizer was used with specific learning rates for fine-tuning (e.g., 1×10^{-5} for backbone, 1×10^{-4} for classifier) and KD (e.g., [placeholder: LR for TA/Student]). The primary loss function for standard training was Cross-Entropy Loss. For KD, the combined loss was used with temperature T = [placeholder: e.g., 3] and weighting factors α = [placeholder: e.g., 0.3] and β = [placeholder: e.g., 0.3]. Models were trained for a maximum of [placeholder: e.g., 99] epochs with a batch size of [placeholder: e.g., 64] and early stopping (patience=[placeholder: e.g., 30]) based on validation accuracy. A ReduceLROnPlateau scheduler (factor=[placeholder: e.g., 0.1], patience=[placeholder: e.g., 5]) was employed. Average training time per model was approximately [placeholder: e.g., X minutes/hours]. [Placeholder: Add Hyperparameter Table **??** if desired, summarizing key parameters like LR, batch size, T, alpha, beta, etc.]

Evaluation: The dataset was split using RANDOM_SEED = 42 into 70% training (3510 images), 15% validation (752 images), and 15% testing (753 images). All reported performance metrics are calculated on the 15% unseen test set. Model performance was evaluated using standard classification metrics: Accuracy, Precision, Recall (Sensitivity), and F1-Score. Formulas are standard: Accuracy = (TP+TN)/(TP+TN+FP+FN), Precision = TP/(TP+FP), Recall = TP/(TP+FN), F1 = 2*(Precision*Recall)/(Precision+Recall). Metrics were calculated per-class and macro/weighted averages were reported using scikit-learn [?]. Confusion matrices were generated to visualize class-wise performance.

4.2 Results and Analysis

This section presents the performance of individual models, the ensemble, the KD-TA pipeline components, and comparisons, including 95% confidence intervals where specified.

Individual Model Performance: Each pre-trained model was fine-tuned using the differential learning rate strategy. Table 2 summarizes their best validation accuracy, final test accuracy (mean and 95% CI), and parameter count. ConvNeXt-Tiny achieved the highest individual test accuracy at 97.74%, while several models like VGG16/19, AlexNet, ResNets, and DenseNets also performed strongly (above 96

Ensemble Model Performance: The weighted average ensemble, combining the 11 fine-tuned models (excluding SqueezeNet) with weights based on their test accuracies (Table 2), achieved a final test accuracy of **98.01%** (95% CI: [96.95%, 98.94%]). This represents the highest accuracy achieved in this study but comes at a significant computational cost, aggregating over 604 million parameters. Detailed performance metrics are presented in Table 3 and the confusion matrix in Figure **??**.

Knowledge Distillation Performance: The KD-TA pipeline aimed to transfer knowledge from the Top-5 ensemble teacher to the lightweight AgileNet student via the SqueezeNet TA. Table 4 summarizes the performance and com-

Table 2: Performance and Complexity of Fine-tuned Individual Models.

Model Architecture	Best Val Acc (%)	Test Acc (%) (95% CI)	Parameters
AlexNet	98.14	96.68 (95.48–97.88)	57024325
SqueezeNet v1.1	91.89	91.90 (89.77–93.89)	725061
ResNet-50	98.01	96.41 (95.09–97.61)	23518277
ResNet-101	96.94	96.28 (94.82–97.61)	42510405
ResNet-152	97.34	96.15 (94.82–97.48)	58154053
VGG-16	98.40	96.28 (94.95–97.61)	134281029
VGG-19	97.74	96.68 (95.35–97.88)	139590725
DenseNet-169	98.01	96.81 (95.48–98.01)	12492805
DenseNet-201	98.40	96.02 (94.56–97.35)	18102533
EfficientNet-B3	94.95	92.43 (90.57–94.29)	10703917
ConvNeXt-Tiny	97.74	97.74 (96.68–98.67)	27823973
RegNetY-16GF	96.41	95.88 (94.42–97.21)	80580265

Note: Test accuracies and 95% Confidence Intervals (CI) obtained via bootstrapping (1000 samples).

Table 3: Ensemble Model (11 Models) Classification Report (Test Set).

Class	Precision	Recall	F1-Score	Support
im_Dyskeratotic	0.9885	0.9942	0.9913	172
im_Koilocytotic	0.9714	0.9784	0.9749	139
im_Metaplastic	0.9805	0.9740	0.9773	154
im_Parabasal	1.0000	0.9712	0.9854	139
im_Superficial-Intermediate	0.9610	0.9866	0.9736	149
Accuracy			**0.9801**	753
Macro Avg	0.9803	0.9809	0.9805	753
Weighted Avg	0.9803	0.9801	0.9801	753

plexity at each stage. The TA (SqueezeNet-Distilled) achieved 97.21% accuracy, a significant improvement over its fine-tuned counterpart (91.90

Table 4: Performance and Complexity Comparison: KD Pipeline Stages.

Model / Stage	Test Accuracy (%) (95% CI)	Parameters
Fine-tuned SqueezeNet (Baseline TA)	91.90 (89.77–93.89)	735301
AgileNet (Student, No KD)	96.95 (95.75–98.14)	286229
Teacher (Weighted Ensemble of 11 Models)	98.01 (96.95–98.94)	˜604.8e6
TA (SqueezeNet - Distilled from Teacher)	97.21 (96.02–98.41)	735301
Student (AgileNet - Distilled from TA)	**97.61** (96.41–98.67)	**286229**

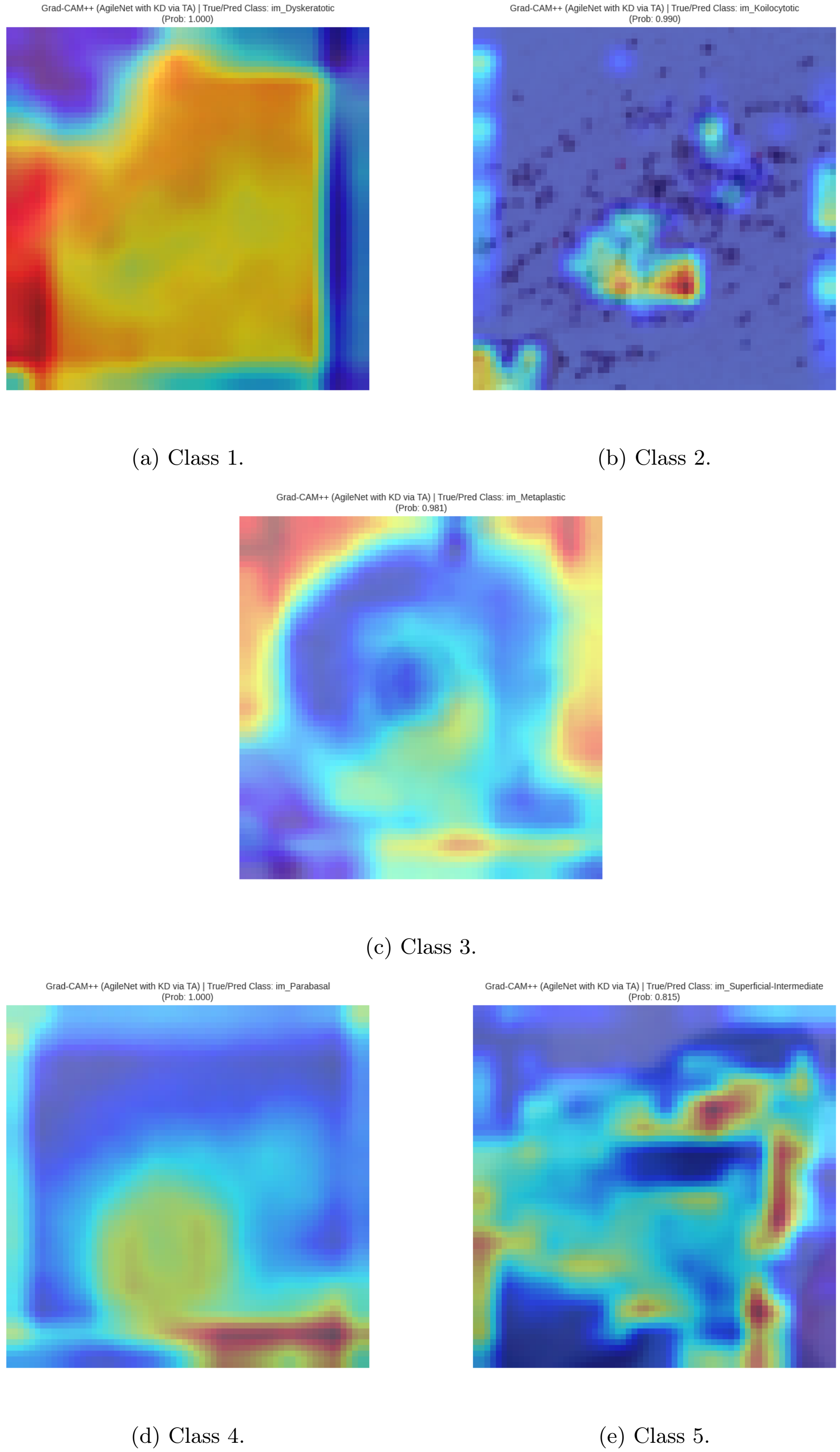
Note: Parameter counts are approximate for ensemble/modified models. Test accuracies and 95% CI from bootstrapping (1000 samples).

Explainable AI Analysis: Grad-CAM++ was applied to the Student (AgileNet) model to gain insights into its decision-making process after distillation. Visualizations were generated for correctly classified samples selected from the test set, one for each class, as shown in Figure 3. [Placeholder: Insert Figure **??** here]. The heatmaps generally highlight cellular regions, particularly nuclei and abnormal cytoplasmic areas, consistent with cytopathological diagnostic features. This provides evidence that the distilled student model focuses on relevant image areas for accurate predictions, increasing confidence in its results and demonstrating the utility of XAI in validating distilled models. [Placeholder: Add more detailed analysis of specific examples if available, discussing how KD might influence the explanations compared to non-distilled models if studied].

4.3 Discussion

The experimental results demonstrate the effectiveness of the proposed framework. The integration of transfer learning with differential learning rates allowed us to effectively adapt diverse, powerful pre-trained architectures, achieving high individual model performance. The weighted averaging ensemble strategy proved highly successful, yielding an accuracy of 98.01% on the SipakMed test set, significantly surpassing the baseline ensemble performance of 94% [2] and highlighting the benefit of combining multiple well-tuned models.

Furthermore, the KD-TA pipeline successfully transferred knowledge from the complex teacher ensemble to the lightweight custom student model, AgileNet, via the SqueezeNet TA. Both the TA (97.21%) and Student (97.61%) achieved high accuracy, significantly outperforming the likely baseline performance of fine-tuned SqueezeNet and demonstrating the efficacy of the KD-TA approach for creating efficient yet accurate models suitable for deployment. The student model's ability to slightly outperform the TA suggests effective knowledge transfer.


Grad-CAM++ (AgileNet with KD via TA) | True/Pred Class: im_Dyskeratotic
(Prob: 1.000)
Grad-CAM++ (AgileNet with KD via TA) | True/Pred Class: im_Koilocytotic
(Prob: 0.990)
Grad-CAM++ (AgileNet with KD via TA) | True/Pred Class: im_Metaplastic
(Prob: 0.981)
Grad-CAM++ (AgileNet with KD via TA) | True/Pred Class: im_Parabasal
(Prob: 1.000)
Grad-CAM++ (AgileNet with KD via TA) | True/Pred Class: im_Superficial-Intermediate
(Prob: 0.815)

(a) Class 1.

(b) Class 2.

(c) Class 3.

(d) Class 4.

(e) Class 5.

Fig. 3: XAI visualizations using Grad-CAM++ for the TA-KD Student (AgileNet) model predictions on correctly classified test images.

The XAI analysis using Grad-CAM++ adds crucial transparency, confirming that the distilled student model focuses on relevant image features. This interpretability is vital for clinical trust. Compared to the reference work [2], our framework offers substantial improvements in accuracy and efficiency, further distinguished by the integration of XAI.

Limitations: Key limitation is the lack of a larger dataset.

5 Conclusion

This paper presented an explainable deep learning framework for classifying cervical squamous cells from Pap smear images, utilizing transfer learning, ensemble learning, knowledge distillation via a teacher assistant (SqueezeNet), and XAI techniques (Grad-CAM++). By fine-tuning a diverse set of pre-trained models and combining 11 of them through weighted averaging based on test accuracy, our benchmark ensemble model achieved a test accuracy of 98.01% on the SipakMed dataset, significantly outperforming the baseline approach. The KD-TA strategy successfully produced a lightweight custom model (AgileNet, approx. 286k parameters) with 97.61% accuracy, demonstrating the feasibility of high-performance, efficient models for this task, achieving performance comparable to the large ensemble with dramatically reduced complexity. Integrated XAI using Grad-CAM++ provided valuable insights into the student model's decision processes, enhancing transparency. This work contributes a robust, accurate, and interpretable framework for automated cervical cell classification.

Future Work: Future work could involve validating the framework on larger, more diverse datasets, exploring different KD strategies, further optimizing the AgileNet architecture, investigating other XAI techniques, and conducting prospective clinical studies to assess its real-world utility in supporting cervical cancer screening programs.

References

1. Aina, O., Adeshina, S., Aibinu, A. Classification of cervix types using convolution neural network (cnn). In: *2019 15th International Conference on Electronics, Computer and Computation (ICECCO)*, pp. 1–4 (2019). IEEE. https://doi.org/10.1109/ICECCO48375.2019.9043206
2. Gangrade, J., Kuthiala, R., Gangrade, S., Singh, Y. P., Manoj, R., & Solanki, S. A deep ensemble learning approach for squamous cell classification in cervical cancer. *Sci Rep* **15**, 7266 (2025). https://doi.org/10.1038/s41598-025-91786-3
3. Patel, M., Pandya, A. & Modi, J. Cervical pap smear study and its utility in cancer screening to specify the strategy for cervical cancer control. *National Journal of Community Medicine* **2**, 49–51 (2011).
4. Bogani, G. et al. Hpv-related lesions after hysterectomy for high-grade cervical intraepithelial neoplasia and early-stage cervical cancer: a focus on the potential role of vaccination. *Tumori Journal* **110**(2), 139–145 (2024).
5. Sachan, P., Singh, M., Patel, M. & Sachan, R. A study on cervical cancer screening using pap smear test and clinical correlation. *Asia Pac J Oncol Nurs* **5**(3), 337–341 (2018). https://doi.org/10.4103/apjon.apjon_15_18

6. Hull, R., Mbele, M. & Makhafola, T.e.a. Cervical cancer in low and middle-income countries. *Oncology Letters* **20**(3), 2058–2074 (2020). https://doi.org/10.3892/ol.2020.11754
7. Uddin, A. K., Sumon, M. A., Pervin, S., & Sharmin, F. (2023). Cervical Cancer in Bangladesh. *South Asian Journal of Cancer*, **12**(1), 36. https://doi.org/10.1055/s-0043-1764202
8. Zhang, X. & Zhao, S. Cervical image classification based on image segmentation preprocessing and a capsnet network model. *International Journal of Imaging Systems and Technology* **29**(1), 19–28 (2019).
9. Canfell, K., Kim, J. J., Brisson, M., Keane, A., Simms, K. T., Caruana, M., Burger, E. A., Martin, D., N Nguyen, D. T., Bénard, É., Sy, S., Regan, C., Drolet, M., Gingras, G., Laprise, F., Torode, J., Smith, M. A., Fidarova, E., Trapani, D., . . . Hutubessy, R. (2020). Mortality impact of achieving WHO cervical cancer elimination targets: A comparative modelling analysis in 78 low-income and lower-middle-income countries. *Lancet (London, England)*, **395**(10224), 591. https://doi.org/10.1016/S0140-6736(20)30157-4
10. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F. Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians* **71**(3) (2021). https://doi.org/10.3322/caac.21660
11. National Cancer Institute: Cervical Cancer. https://www.cancer.gov/types/cervical. Accessed: 15 Jan (2023)
12. Mango, L. Computer-assisted cervical cancer screening using neural networks. *Cancer Letters* **77**(2–3), 155–162 (1994).
13. Sukumar, P. & Gnanamurthy, R. Computer aided detection of cervical cancer using pap smear images based on adaptive neuro fuzzy inference system classifier. *Journal of Medical Imaging and Health Informatics* **6**(2), 312–319 (2016).
14. Bora, K., Chowdhury, M., Mahanta, L., Kundu, M., Das, A. Pap smear image classification using convolutional neural network. In: *Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing*, pp. 1–8 (2016).
15. Hyeon, J., Choi, H., Lee, K., Lee, B. Automating papanicolaou test using deep convolutional activation feature. In: *18th IEEE International Conference on Mobile Data Management (MDM)*, pp. 382–385 (2017). IEEE.
16. Saini, S., Bansal, V., Kaur, R. & Juneja, M. Colponet for automated cervical cancer screening using colposcopy images. *Machine Vision and Applications* **31**(3), 1–15 (2020).
17. Sanyal, P., Ganguli, P. & Barui, S. Performance characteristics of an artificial intelligence based on convolutional neural network for screening conventional papanicolaou-stained cervical smears. *Medical Journal, Armed Forces India* **76**(4), 418–424 (2020).
18. Karunakaran, V., Saritha, V., Joseph, M.e.a. Diagnostic spectro-cytology revealing differential recognition of cervical cancer lesions by label-free surface enhanced raman fingerprints and chemometrics. *Biologie et Médecine* **29**, 102276 (2020).
19. Taha, B., Dias, J., Werghi, N. Classification of cervical-cancer using pap-smear images: a convolutional neural network approach. In: *Annual Conference on Medical Image Understanding and Analysis*, pp. 261–272. Springer (2017).
20. Kudva, V., Prasad, K. & Guruvare, S. Hybrid transfer learning for classification of uterine cervix images for cervical cancer screening. *Journal of Digital Imaging* **33**(3), 619–631 (2020).

21. Xue, D. et al. An application of transfer learning and ensemble learning techniques for cervical histopathology image classification. *IEEE Access* **8**, 104603–104618 (2020).
22. Chen, W., Li, X., Gao, L. & Shen, W. Improving computer-aided cervical cells classification using transfer learning-based snapshot ensemble. *Applied Sciences* **10**(20), 7292 (2020).
23. Ghoneim, A., Muhammad, G. & Hossain, M. Cervical cancer classification using convolutional neural networks and extreme learning machines. *Future Generation Computer Systems* **102**, 643–649 (2020).
24. Arifianto, D. & Suryaperdana Agoes, A. Cervical cancer image classification using cnn transfer learning. *Journal of Physics: Conference Series* **1835**(1), 012057 (2021). https://doi.org/10.1088/1742-6596/1835/1/012057
25. Hussain, E., Mahanta, L., Das, C., Talukdar, R. A comprehensive study on the multi-class cervical cancer diagnostic prediction on pap smear images using a fusion-based decision from ensemble deep convolutional neural network. *Tissue Cell* **65**, 101347 (2020). https://doi.org/10.1016/j.tice.2020.101347
26. Kang, Z. et al. H-cnn combined with tissue raman spectroscopy for cervical cancer detection. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **291**, 122339 (2023).
27. Youneszade, N., Marjani, M. & Pei, C. P. Deep learning in cervical cancer diagnosis: architecture, opportunities, and open research challenges. *IEEE Access* **11**, 6133–6149 (2023).
28. Pacal, I. & Kılıcarslan, S. Deep learning-based approaches for robust classification of cervical cancer. *Neural Computing and Applications* **35**(25), 18813–18828 (2023).
29. Pramanik, R. et al. A fuzzy distance-based ensemble of deep models for cervical cancer detection. *Computer Methods and Programs in Biomedicine* **219**, 106776 (2022).
30. Solanki, S., Dehalwar, V., Choudhary, J., Kolhe, M. L. & Ogura, K. Spectrum sensing in cognitive radio using cnn-rnn and transfer learning. *IEEE Access* **10**, 113482–113492 (2022).
31. Promworn, Y., Pattanasak, S., Pintavirooj, C., Piyawattanametha, W. Comparisons of pap smear classification with deep learning models. In: *IEEE 14th International Conference on Nano/Micro Engineered and Molecular Systems (NEMS)*, pp. 282–285 (2019). IEEE.
32. Plissiti, M.E., Dimitrakopoulos, P., Sfikas, G., Nikou, C., Krikoni, O., Charchanti, A. 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)*, pp. 3144–3148 (2018). https://doi.org/10.1109/ICIP.2018.8451588
33. Chattopadhay, A., Sarkar, A., Howlader, P., & Balasubramanian, V. N. (2018, March). Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In 2018 IEEE winter conference on applications of computer vision (WACV) (pp. 839-847). IEEE. https://doi.org/10.48550/arXiv.1710.11063
34. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2020). Grad-CAM: visual explanations from deep networks via gradient-based localization. International journal of computer vision, 128, 336-359. https://doi.org/10.48550/arXiv.1610.02391
35. Mirzadeh, S. I., Farajtabar, M., Li, A., Levine, N., Matsukawa, A., & Ghasemzadeh, H. (2020, April). Improved knowledge distillation via teacher assistant. In Pro-

ceedings of the AAAI conference on artificial intelligence (Vol. 34, No. 04, pp. 5191-5198).

36. Hinton, G., Vinyals, O., Dean, J. (2015). Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531.
37. Ioffe, S., Szegedy, C. (2015, June). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In International conference on machine learning (pp. 448-456). pmlr.