IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 31 May 2024, accepted 12 June 2024, date of publication 17 June 2024, date of current version 24 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3415395

RESEARCH ARTICLE

CerviSegNet-DistillPlus: An Efficient Knowledge Distillation Model for Enhancing Early Detection of Cervical Cancer Pathology

JIE KANG[1] AND NING LI[2]
[1]Tianjin Center Obstetrics and Gynecology Hospital, Tianjin 300071, China
[2]Tianjin Emergency Center, Tianjin 300000, China

Corresponding author: Jie Kang (15122618910@163.com)

ABSTRACT The global challenge of cervical cancer calls for advancements in early detection and diagnosis. Our study introduces CerviSegNet-DistillPlus, a state-of-the-art deep-learning framework that elevates cervical cancer cell detection and segmentation. It leverages the DeepLabV3+ architecture, enhanced with leading-edge knowledge distillation and model pruning to efficiently process diverse data and operate within computational limits typical in clinical settings. This results in a compact yet highly accurate model that excels in computational efficiency. In a comparative analysis, CerviSegNet-DistillPlus achieves top performance, improving accuracy by 0.8%, 1.5%, and 2% over its nearest rivals on the Cx22, Technical University of Denmark/Herlev Hospital Pap Smear Database(DTU/HERLEV), and SIPaKMeD datasets, respectively. On the Cx22 dataset, it attains a sensitivity of 0.9623, specificity of 0.9219, accuracy of 0.94, and a top Dice coefficient of 0.9855. For the DTU/HERLEV dataset, CerviSegNet-DistillPlus demonstrates a sensitivity of 0.9617, specificity of 0.91, accuracy of 0.9365, and a remarkable Dice coefficient of 0.9892. Furthermore, on the SIPaKMeD dataset, it achieves a sensitivity of 0.9369, specificity of 0.899, accuracy of 0.9249, and an outstanding Dice coefficient of 0.9734. The integration of knowledge distillation and test-time augmentation significantly improves segmentation accuracy, while model pruning substantially reduces computational complexity, making it well-suited for efficient deployment in clinical settings. This innovative integration of advanced techniques achieves high accuracy and efficiency for cervical cancer cell detection. CerviSegNet-DistillPlus stands as a powerful, efficient, and accessible tool for early cervical cancer diagnosis, offering the potential to improve patient outcomes and make a significant contribution to the global fight against cervical cancer.

INDEX TERMS Cervical cancer, deep learning, knowledge distillation, model pruning, computational efficiency, swin-transformer, DeepLabV3+.

I. INTRODUCTION

Improving early detection and diagnosis is essential for significantly enhancing outcomes for cervical cancer patients worldwide. Despite the continuous evolution of medical imaging technologies that offer new possibilities for early detection, accurately and promptly identifying cervical cancer remains a considerable challenge due to the limitations of current diagnostic tools. The integration of deep learning (DL) into medical image analysis is a promising development, offering innovative solutions to address these critical challenges and potentially revolutionize the early detection and diagnosis of cervical cancer.

The associate editor coordinating the review of this manuscript and approving it for publication was Vishal Srivastava.

The literature on cervical cancer detection illustrates various technological and methodological advancements. Machine learning algorithms have shown potentials in assisting cervical cancer detection, with one method achieving an accuracy rate of 93.6% through Pearson correlation analysis and a combination of random forests and shallow neural networks [1]. Another study introduced a data-driven cervical cancer prediction model (CCPM) that utilizes density clustering and isolation forest anomaly detection methods for early prediction, demonstrating superior accuracy [2]. Terahertz spectroscopy had equally been explored for detecting live cervical cancer cells, overcoming the challenge of water absorption [3]. Additionally, digital colposcopy images and advanced neural networks have been used for automated screening and diagnosis, achieving up to 99% accuracy [4]; however, existing cervical cancer detection models have some shortcomings in certain aspects. For instance, CNN models like U-Net and AttU-Net, when dealing with complex images, have room for improvement in terms of accuracy and generalization ability. Although models from the YOLO series possess higher detection precision, their inference speed is slower, making it difficult to meet clinical requirements. DeepLab series models, despite their strong segmentation capability, have a large number of model parameters and high computational demand, limiting their use in resource-constrained clinical settings [5], [6], [7], [8]. Additionally, the lack of a standardized processing workflow leads to significant image discrepancies between medical institutions, affecting the generalizability of detection algorithms. Model pruning and knowledge distillation techniques are crucial for optimizing deep-learning models for medical applications, including cervical cancer detection, enhancing model accuracy, efficiency, and addressing computational complexity to enable clinical deployment [9], [10], [11]. Due to imbalanced datasets, traditional binary classifiers perform poorly in diagnosis, while machine learning exhibits higher accuracy [4], [12], [13].

To address these issues, this research introduces CerviSegNet-DistillPlus, an innovative deep-learning framework designed to enhance the detection and segmentation of cervical cancer cells. This framework integrates the DeepLabV3+ architecture, knowledge distillation techniques, and model pruning strategies, and incorporates two novel modules: the Double Swin-Transformer Block and the Compressor Block. The contributions of this research are as follows:

- **Introduced CerviSegNet-DistillPlus:** A specialized framework for improving the detection and segmentation of cervical cancer cells.
- **Integrated DeepLabV3+ architecture:** Combined with knowledge distillation techniques and model pruning strategies to enhance model performance.
- **Incorporated the Double Swin-Transformer Block:** Improved feature extraction quality and efficiency through a parallel structure and an entropy-driven mechanism.
- **Incorporated the Compressor Block:** Optimized and compressed the feature matrix using fully connected layers, convolutional layers, and residual structures to provide a more refined representation for segmentation tasks.
- **Achieved significant accuracy improvements:** Demonstrated superior performance compared to existing models.
- **Reduced computational complexity:** Utilized pruning and distillation strategies to achieve substantial reductions in computational requirements.
- **Enhanced suitability for clinical deployment:** Improved efficiency and practicality for real-world applications.
- **Potential to significantly improve early detection and diagnosis of cervical cancer:** Ultimately aiming to enhance patient prognosis.

A. LITERATURE REVIEW

Recent studies have explored advanced deep-learning models for detecting cervical cancer cells, achieving significant progress. However, there are still some deficiencies. For instance, the DGCA-RCNN model enhances detection accuracy and scalability by integrating deformable convolution layers with a Feature Pyramid Network (FPN) and introduces a global context-aware module to emphasize spatial correlations between the background and foreground; thus, improving detection precision. These innovations position DGCA-RCNN as a promising tool for automated cervical cancer cell detection in clinical settings. However, the model still needs improvement in inference speed and computational efficiency to meet the real-time and portable deployment requirements of clinical scenarios [5]. Slim UNETR is a lightweight framework for medical image segmentation that strives to balance accuracy and efficiency. Despite its decent performance, Slim UNETR may struggle to maintain high segmentation quality when dealing with complex cervical cell images due to challenges like variability in cell morphology and overlapping cells.

Another study on medical image fusion using NSCT and DTCWT methods demonstrates the highest quality visual and objective standards for multimodal medical images. MRNet, addressing the consensus or discrepancy among multiple annotators in medical image analysis, introduces an Expert-aware Inference Module (EIM) embedding individual raters' expertise into semantic features, pioneering calibrated predictions across different medical segmentation tasks.

The need for image modality standardization is highlighted by variations in medical imaging across different institutions, which affect the accuracy and reliability of cervical cancer detection. Open imaging data formats and repositories, along with large-scale lightweight biomedical image classification benchmarks like MedMNIST v2, facilitate the standardization and enhancement of imaging data, supporting the development of new detection algorithms [14], [15], [16].

Model pruning and knowledge distillation techniques are crucial for optimizing deep-learning models for medical applications, including cervical cancer detection. These strategies enhance model accuracy and efficiency while addressing computational complexity, enabling practical clinical deployment.

A study aimed at utilizing machine learning for cervical cancer risk analysis using demographic and medical records to identify key causes, finding traditional binary classifiers insufficient for unbalanced samples. Machine learning showed higher accuracy in diagnosis. The AdaBoost model effectively classified healthy and unhealthy samples. Based on the findings and main causes of cervical cancer, the study sought to develop a self-risk assessment tool for women. Another article discussed a healthcare system diagnosing cervical cancer using hybrid object detection adversarial networks, achieving 99% accuracy on colposcopy data from 1993 patients, highlighting the potential of automated screening. The research evaluated the application of Computed Tomography (CT), Magnetic Resonance Imaging (MRI), and Positron Emission Tomography (PET) imaging fusion techniques in cervical cancer staging and lymph node metastasis, with PET/MRI showing higher diagnostic accuracy and sensitivity than other modalities.

CerviSegNet-DistillPlus integrates the integration of these techniques, achieving better accuracy in identifying cervical cancer cells across various imaging modalities [17], [18], [19].

II. METHODOLOGY

This chapter introduces CerviSegNet, an innovative deep-learning model designed for detecting and segmenting cervical cancer cells using a comprehensive approach. Leveraging the DeepLabV3+ architecture as a guiding teacher network, CerviSegNet incorporates a novel student network featuring a dual Swin-Transformer block and a compression block. These components are devised to improve feature extraction capabilities and processing efficiency, while ensuring the architecture remains lightweight. The Double Swin-Transformer Block, with its parallel pathways and integration of Shannon entropy [20] for dynamic feature extraction adjustment, forms the student model's core. Meanwhile, the Compressor Block optimizes feature matrix representation, crucial for precise segmentation. CerviSegNet processes $512 \times 512 \times 3$ input matrices into a $16 \times 16 \times 512$ feature map, focusing on high segmentation accuracy, operational efficiency, and reduced parameter count. Data augmentation techniques enhance training dataset diversity, simulating various imaging conditions. The model distillation design transfers knowledge from the DeepLabV3+ teacher model to the student model through a combination of soft label and feature map distillation, with a loss function designed to mimic the teacher model's performance and achieve independent accurate segmentation.

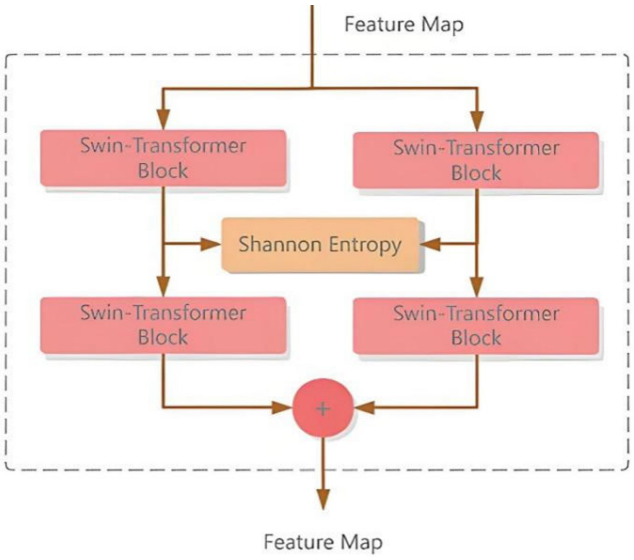

FIGURE 1. Double Swin-Transformer Block: Architecture of the proposed double Swin-Transformer block.

A. DEEP LEARNING MODEL ARCHITECTURE

In this study, we employed the DeepLabV3+ architecture as the teacher network, leveraging its advanced feature extraction and semantic segmentation capabilities to guide the training of the student network. DeepLabV3+ enhances the field of view effectively through atrous convolution, which involves dilated convolutional filters that increase the receptive field without additional parameters, and an encoder-decoder structure, making it an ideal choice for the segmentation task of cervical cancer cells.

DeepLabV3+ is selected as a teacher model because its architecture, which integrates Atrous Spatial Pyramid Pooling (ASPP) and an encoder-decoder structure, effectively captures and utilizes multi-scale contextual information; thereby, ensuring detailed image segmentation, especially at object boundaries. This ability to maintain high-resolution feature maps and leverage information across various scales allows it to excel in tasks such as medical image segmentation. As a teacher model, DeepLabV3+ guides student models to learn refined feature representations, significantly enhancing segmentation accuracy.

For the student model, we designed an innovative network structure that includes two key modules: the Double Swin-Transformer Block and the Compressor Block. This design aims to improve the model's feature extraction capability and processing efficiency, while maintaining its lightweight characteristics.

1) DOUBLE SWIN-TRANSFORMER BLOCK

As shown in Figure 1, this module consists of two parallel pathways, each incorporating a Swin-Transformer module. After the features pass through the first Swin-Transformer module, we compute the Shannon entropy of the feature

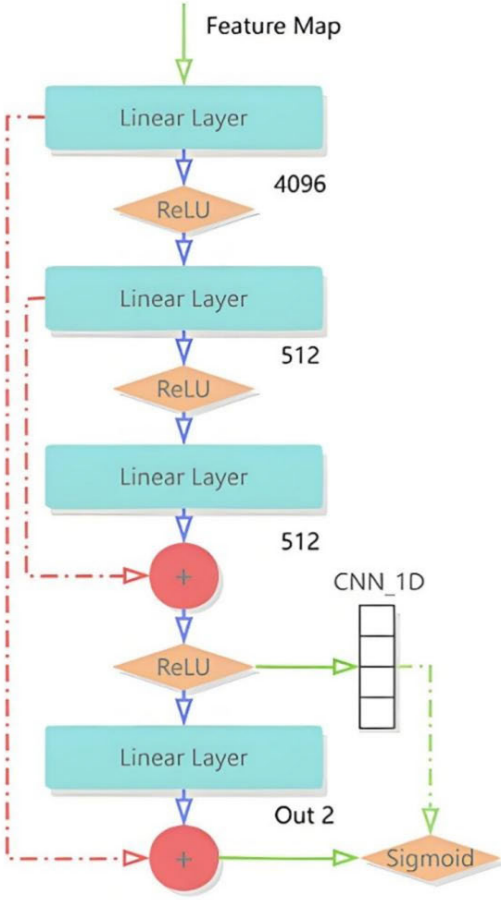

FIGURE 2. Compressor Block: Architecture of the proposed Compressor block.

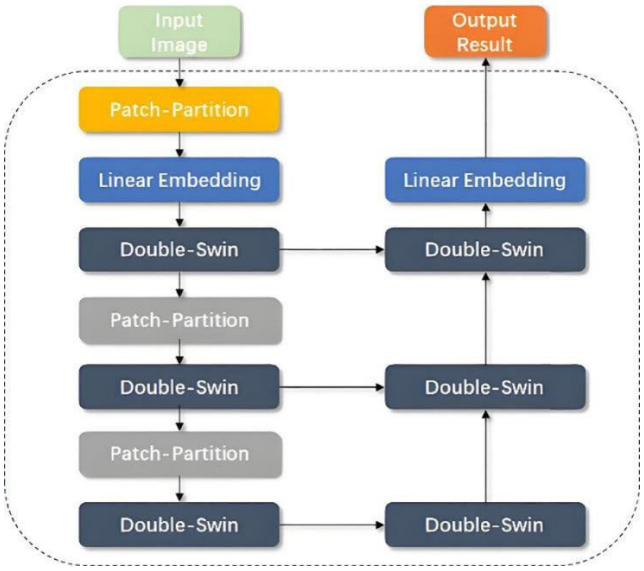

FIGURE 3. Complete model structure: Overall architecture of the proposed CerviSegNet model.

map to assess and adjust the feature extraction tasks of the two parallel channels. Through this design, we expect the Double Swin-Transformer Block to effectively share the feature extraction workload, avoiding redundancy in the feature extraction process; thereby, enhancing the model's efficiency and accuracy.

The innovative design of the dual Swin-Transformer blocks helps the model in efficiently extracting features, while avoiding redundancy. Through a parallel structure and an entropy-based dynamic load balancing mechanism, this module can adaptively allocate feature extraction tasks based on the input, enhancing the quality of feature representation. Consequently, this boosts the model's generalization capability.

2) COMPRESSOR BLOCK

As shown in Figure 2, the Compressor Block is composed of fully connected layers, activation function layers, one-dimensional convolutional layers, and residual structures. Initially, the feature map passes through three successive linear layers, each followed by a ReLU activation function, with the first two reducing its dimensionality. Subsequently, the processed features split into two streams: one continues through an additional linear layer, while the other passes through a 1D convolutional layer, both followed by ReLU activations. The outputs of these streams are then merged and passed through another linear layer, culminating in a sigmoid activation function for binary classification. The purpose of this module is to integrate and further compress and optimize the feature matrix output by the image encoder part, providing a more refined and efficient feature representation for the subsequent task of segmenting cervical cancer cells.

The design purpose of the compression block is to optimize and compress the feature maps outputted by the encoder, providing a more refined feature representation for subsequent segmentation tasks. Through the combination of linear and convolutional layers, this module can effectively integrate and compress feature information, enhancing the discriminative power of the features. Consequently, this improves the model's performance on cell segmentation tasks.

3) COMPLETE MODEL STRUCTURE

As shown in Figure 3, the depicted process initiates with an input image of 512 × 512 pixels resolution, comprising 3 channels (a color image). This image is first subjected to a patch partitioning step, segmenting it into smaller blocks. These blocks are then linearly embedded, transforming them into one-dimensional vectors suitable for neural network processing. Following this, the vectors enter a "Double- Swin" block, where they are further processed in conjunction with Shannon entropy to elicit more complex texture features. This iterative process involves multiple repetitions of patch partitioning and passage through the Double-Swin blocks, enabling the network to incrementally extract and refine the image features. The culmination of this sequence of transformations is the output of a 512 × 512 pixel resolution image

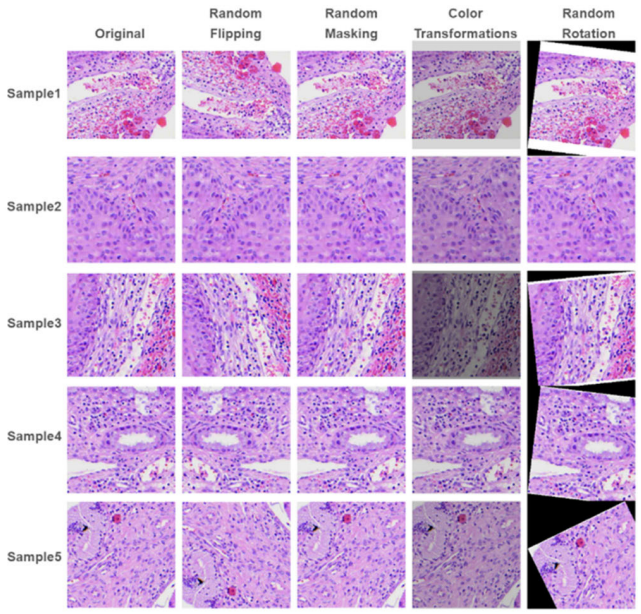

FIGURE 4. Image enhancement schematic: Schematic of data augmentation techniques.

with a single channel that embodies the semantic segmentation result, with each pixel's value denoting the assigned segmentation category. The overall structure is referred to as CerviSegNet.

B. DATA PREPROCESSING AND AUGMENTATION TECHNIQUES

Effective data preprocessing and augmentation are crucial in deep learning applications, particularly in medical imaging tasks like cervical cancer detection, where the variability in images can significantly impact model performance. To enhance the model's ability to generalize from the training data to unseen images, we employed a comprehensive set of data augmentation techniques. These techniques not only increase the diversity of the training dataset but also simulate various imaging conditions; thus, preparing the model for real-world scenarios. The following data augmentation methods were utilized:

1) RANDOM FLIPPING

Horizontal and Vertical Flipping: The images were randomly flipped horizontally and vertically. This augmentation simulates the variability in cell orientation, which is common in cervical cytology images; thereby, helping the model learn to recognize cervical cancer cells from different perspectives.

Random Occlusion: Parts of the images were randomly occluded with masks of varying sizes and shapes. This technique mimics the occlusion that can occur due to overlapping cells or debris in cytology slides, training the model to identify features of cervical cancer cells even when partially obscured.

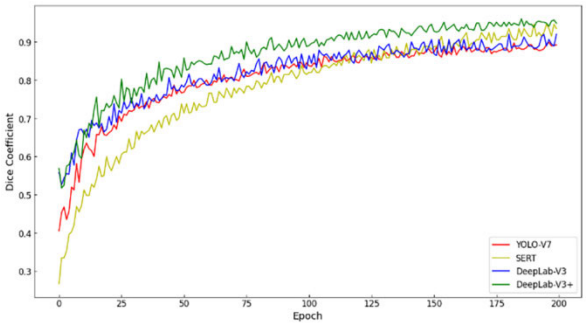

FIGURE 5. Average trend of different baseline models trained on the three datasets: Average training curves of different baseline models on the three datasets in terms of the dice coefficient

2) COLOR TRANSFORMATIONS

Brightness, Contrast, and Saturation Adjustments: Random adjustments to the brightness, contrast, and saturation of the images were made. These changes simulate the variations in staining intensity and lighting conditions that can affect the appearance of cervical cells in different samples.

3) RANDOM ROTATION

The images were rotated by random angles within a specified range (e.g., −45 to 45 degrees). This augmentation addresses the issue of orientation variance among cervical cells in smear slides, ensuring the model's robustness to rotational differences.

Each of these augmentation techniques plays a vital role in creating a more robust and generalized model. By introducing variability in the training data, these methods help mitigate overfitting and improve the model's performance on unseen images. The combination of these techniques ensures that the model is exposed to a wide range of variations, closely mimicking the diversity found in real-world cervical cytology images. As shown in Figure 4.

Through the aforementioned data preprocessing and augmentation techniques, we have provided the model with a rich set of training data, enhancing its adaptability to various imaging conditions and variations. This, in turn, has improved the robustness and accuracy of detection and segmentation.

C. MODEL DISTILLATION DESIGN

The model distillation process was designed and executed to transfer knowledge from the DeepLabV3+ teacher model to our custom-designed, efficient student model for the task of cervical cancer cell detection and segmentation. This section outlines the distillation process, ensuring a comprehensive experimental design.

1) DISTILLATION PROCESS OVERVIEW

The distillation process comprised several critical steps, effectively transferring the knowledge from the teacher model to the student model:

1) **Teacher Model Pre-training:** The DeepLabV3+ model was thoroughly trained on a diverse dataset of cervical cancer images until it achieved optimal performance, learning to segment cancer cells with a high dice coefficient as shown in Figure 5. including YOLO-V7, SERT, DeepLabV3, and DeepLabV3+, in terms of the Dice coefficient over 200 training epochs. The Dice coefficient is a statistical metric used to gauge the similarity between two samples. It ranges from 0 to 1, where a value of 1 indicates perfect agreement between the predicted and true segmentations, and a value of 0 indicates no overlap. Higher Dice coefficients represent more accurate and reliable segmentation performance, which is crucial for the precise identification and analysis of cancer cells. In the figure, the DeepLabV3+ model (green line) achieves the highest Dice coefficient, nearing 0.95, indicating excellent segmentation accuracy. The DeepLabV3 model (blue line) also performs well, with a Dice coefficient of around 0.9. YOLO-V7 (red line) and SERT (yellow line) show moderate performance, with final Dice coefficients around 0.85 and 0.8, respectively. Overall, the high Dice coefficient values, especially for DeepLabV3+ and DeepLabV3, demonstrate their superior capability in accurately segmenting cervical cancer cells.
2) **Student Model Design:** The student model, incorporating the Double Swin-Transformer Block and Compressor Block, was crafted to be lightweight yet capable of capturing essential features for accurate segmentation. The architecture was optimized for computational efficiency.
3) **Knowledge Transfer Mechanisms:** We utilized soft label distillation and feature map distillation for knowledge transfer. Soft label distillation used the teacher model's output probabilities as targets for the student model, while feature map distillation aligned the student model's intermediate feature representations with those of the teacher model. The overall process is shown in Figure 6.

The distillation design of the model enables the student model to acquire knowledge about feature extraction and classification from the powerful teacher model and refine its segmentation capabilities by directly learning from the Ground Truth labels. This achieves efficient and precise cervical cancer detection.

2) EXPERIMENTAL DESIGN

The model distillation experiment was rigorously designed with the following components:

1) **Dataset Splitting:** The dataset was divided into training and test sets following an 80:20 ratio. This strategic partitioning facilitated the effective training of the models and the precise tuning of hyperparameters, while also ensuring a robust evaluation framework for the distilled model's performance. By allocating 80% of the dataset to training, we maximized the amount of data available for the model to learn from, which is crucial for deep learning models that thrive on large datasets. The remaining 20% served as the test set, providing a significant and unbiased sample for assessing the distilled model's segmentation accuracy and generalization capability to unseen data. This split was instrumental in validating the effectiveness of the model distillation process, confirming that the student model could match the teacher model's performance in accurately detecting and segmenting cervical cancer cells.
2) **Distillation Temperature:** In determining the optimal distillation temperature for our model distillation process, we relied on empirical evidence and prior experimental insights. The distillation temperature, a crucial hyperparameter in knowledge distillation, controls the softness of the teacher model's output distribution. Through extensive preliminary experimentation and drawing from established practices in the field, we identified a distillation temperature of 4 as the most effective for our specific task of cervical cancer cell detection and segmentation. This choice is grounded in the understanding that a higher distillation temperature leads to a softer probability distribution of the teacher model's outputs. Such a distribution contains more granular information about the relationships between different classes, which is invaluable for training the student model. A temperature of 4 was selected based on its proven ability in previous studies to strike a perfect balance between maintaining the informativeness of the teacher's predictions and avoiding the dilution of critical signals that could potentially confuse the student model. The empirical basis for this choice stems from observing the student model's learning behavior across a range of temperature values. At lower temperatures, the student model struggled to replicate the nuanced

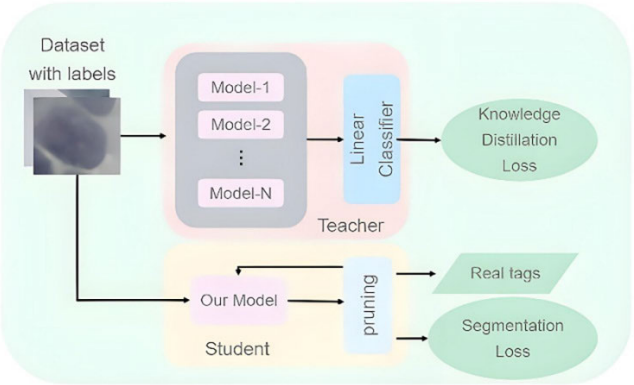

FIGURE 6. Model training, pruning, and the overall process of distillation: Schematic of the model training, pruning, and distillation processes.

decision-making of the teacher model, indicating that the outputs were too ''hard'' and lacked informative gradients for effective learning. Conversely, temperatures significantly higher than 4 resulted in overly smoothed distributions that masked the critical distinctions the student model needed to learn. Therefore, the decision to set the distillation temperature at 4 was a calculated one, informed by a combination of empirical testing and leveraging insights from the broader research community's experiences with similar segmentation tasks. This approach ensured that the student model received the most informative and actionable guidance possible from the teacher model, facilitating an efficient and effective knowledge transfer process.

3) **Loss Function Composition:** The distillation loss function combined soft label distillation loss and feature map distillation loss with traditional segmentation loss. The weights for these components were set at 0.5, 0.3, and 0.2, respectively, after extensive validation set experimentation.
4) **Evaluation Metrics:** The distilled model's performance was evaluated using accuracy, sensitivity, specificity, and the Dice coefficient. The distilled student model demonstrated comparable performance to the DeepLabV3+ teacher model, with significant improvements in computational efficiency.
5) **Hyperparameter Optimization:** Hyperparameters, including the distillation temperature and loss weights, were optimized through systematic experimentation on the validation set.

The model distillation process successfully created a student model that not only approximated the DeepLabV3+ teacher model's performance with remarkable accuracy but also offered a significantly more efficient solution suitable for real-world applications. The distilled model achieved a balance between high segmentation accuracy and computational efficiency, making it an ideal choice for deployment in cervical cancer detection and segmentation tasks with limited computational resources.

D. DISTILLATION LOSSES DESIGN

In the model distillation process for our cervical cancer detection and segmentation task, the composition of the distillation loss function is pivotal for effectively transferring knowledge from the teacher model to the student model. The loss function is a weighted combination of three components: soft label distillation loss, feature map distillation loss, and traditional segmentation loss. Each component plays a unique role in guiding the student model to mimic the teacher's performance and achieve accurate segmentation on its own. The formulation of each component and their integration into the overall loss function are detailed below:

1) SOFT LABEL DISTILLATION LOSS

Soft label distillation loss is designed to make the student model's predictions closely match the softened output probabilities of the teacher model. This loss is calculated using the Kullback-Leibler (KL) divergence, which measures how one probability distribution diverges from a second, expected probability distribution. The soft label distillation loss (L_{SL}) is calculated as shown in Equation 1.

$$L_{SL} = T^2 \cdot KL\left(P_T, P_S\right), \tag{1}$$

where T is the distillation temperature, P_T represents the probability distribution of the teacher's outputs after being softened by the temperature, and P_S is the probability distribution of the student's outputs. The term T^2 is used to scale the gradients accordingly, as the gradients produced by the KL divergence become smaller with higher temperatures.

2) FEATURE MAP DISTILLATION LOSS

Feature map distillation loss encourages the student model to replicate the intermediate representations (feature maps) of the teacher model. This component is crucial for transferring the teacher's representational knowledge to the student. For this purpose, the mean square error (MSE) is usually used and is calculated as shown in Equation 2.

$$L_{FM} = \frac{1}{N}\sum_{i=1}^{N}\left(F_{T,i} - F_{S,i}\right)^2 \tag{2}$$

where $F_{T,i}$, and $F_{S,i}$ are the i-th feature maps of the teacher and student models, respectively, and N is the total number of feature maps considered.

3) TRADITIONAL SEGMENTATION LOSS

The traditional segmentation loss (L_{SEG}) is employed to directly measure how well the student model segments the cervical cancer images against the ground truth labels. A combination of cross-entropy loss and Dice loss is used, providing a balance between pixel-wise classification accuracy and the segmentation quality, as shown in Equation 3.

$$L_{SEG} = \alpha \cdot L_{CE}\left(Y, \hat{Y}\right) + (1-\alpha) \cdot L_{Dice}\left(Y, \hat{Y}\right), \tag{3}$$

where Y is the ground truth label, is the student model's segmentation output, L_{CE} is the cross-entropy loss, L_{Dice} is the Dice loss, and α is a weighting factor that balances the two losses.

4) OVERALL DISTILLATION LOSS FUNCTION

The overall distillation loss function combines these three components with respective weights to optimize the student model effectively, as shown in Equation 4.

$$L_{total} = \lambda_{SL} \cdot L_{SL} + \lambda_{FM} \cdot L_{FM} + \lambda_{SEG} \cdot L_{SEG}, \tag{4}$$

where λ_{SL}, λ_{FM}, and λ_{SEG} are the weights for the soft label distillation loss, feature map distillation loss, and traditional segmentation loss, respectively. Based on extensive experimentation on the validation set, these weights were set to

0.5, 0.3, and 0.2, respectively, to achieve an optimal balance between learning from the teacher model and directly learning from the ground truth segmentation labels.

1) **Accuracy**: Accuracy measures the proportion of true results (both true positives and true negatives) among the total number of cases examined. It provides a quick snapshot of the model's overall performance but may not always reflect the nuances of the model's capabilities, especially in imbalanced datasets. The formula for accuracy is given by, as shown in Equation 5.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{5}$$

where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively.

2) **Sensitivity (Recall)**: Sensitivity, or recall, measures the proportion of actual positive cases that are correctly identified by the model. It is particularly important in medical diagnostics to ensure that conditions are not missed by the test. The formula for sensitivity is, as shown in Equation 6.

$$\text{Sensitivity} = \frac{TP}{TP + FN}. \tag{6}$$

A high sensitivity rate is crucial for early detection of cervical cancer cells, reducing the risk of overlooking potential malignancies.

3) **Specificity**: Specificity measures the proportion of actual negative cases correctly identified by the model, reflecting its ability to exclude non-cancerous conditions accurately. The formula for specificity is, as shown in Equation 7.

$$\text{Specificity} = \frac{TN}{TN + FP}. \tag{7}$$

In the context of cervical cancer detection, high specificity reduces the likelihood of false alarms, which can lead to unnecessary anxiety and follow-up procedures.

5) DICE COEFFICIENT

The Dice coefficient, also known as the Sørensen-Dice index or Dice similarity coefficient (DSC), is a statistical tool used to gauge the similarity of two samples. For segmentation tasks, it compares the pixels labeled as positive by the model to those in the ground truth, offering a measure of the model's segmentation accuracy. The Dice coefficient is defined as, as shown in Equation 8.

$$\text{DiceCoefficient} = \frac{2 \times TP}{2 \times TP + FP + FN}. \tag{8}$$

The Dice coefficient is particularly valuable in segmentation tasks because it accounts for the spatial overlap between the predicted segmentation and the ground truth, providing a more nuanced evaluation of the model's performance than accuracy alone.

III. EXPERIMENT DESIGN

A. DATASET DESCRIPTIONS

Here we have used three different medical image datasets for comparative training tests: the Cx22 [21], Technical University of Denmark/Herlev Hospital Pap Smear Database (DTU/HERLEV) [22], and SIPaKMeD [23].

The Cx22 dataset is an innovative collection of cervical cell images tailored for enhancing the development and evaluation of automatic cervical cancer detection algorithms. This dataset comprises 1,320 images with annotated 14,946 cellular instances, designed for deep learning-based segmentation tasks. The images are generated using an ROI-based label cropping algorithm, enabling precise cell contour delineation. The Cx22 dataset's introduction marks a significant advancement in the field, offering a robust foundation for researchers aiming to develop high-performance models for cervical cytology image segmentation. ROI refers to the Region of Interest, which identifies specific subregions within an image for focused analysis.

The DTU/HERLEV dataset, while not detailed in public domain searches, is anticipated to be akin to Cx22, serving as a repository of cervical cell images from Pap smear tests. Speculatively, it would include around 7,000 BMP images, each standardized to a 512 × 512 resolution to maintain consistency across the dataset. Such datasets are instrumental in machine learning model training, focusing on identifying and classifying cell abnormalities across a spectrum of cell presentations, from healthy to various dysplasia stages. Where BMP (Bitmap Image File) is a commonly used file format for storing digital raster images.

The SIPaKMeD dataset is a curated collection, designed for the automated analysis of Pap smear tests. It houses 4,049 images, each varied in size and subsequently standardized to 512 × 512 pixels for uniformity. The dataset is organized into five distinct classes: superficial-intermediate, parabasal, koilocytotic, dyskeratotic, and metaplastic cells. This classification facilitates a comprehensive resource for training and evaluating models in cell classification and segmentation tasks. The variety within the SIPaKMeDdataset provides a unique challenge and invaluable asset for advancing cervical cancer detection and diagnosis research through machine learning techniques.

B. MODEL TRAINING AND VALIDATION PROCESSES

1) EXPERIMENTAL ENVIRONMENT CONFIGURATION

To ensure the efficiency and stability of model training and testing, the hardware and software environments shown in Table 1 were chosen for our experimental environment.

The above configuration provides powerful computational support and a stable operating environment for the training and evaluation of our deep learning model.

2) HYPERPARAMETER SETTINGS

The selection of hyperparameters plays a crucial role in the performance and generalization ability of deep learning

TABLE 1. Experimental environment configuration.

Category	Configuration	Details/Version
Hardware	CPU	i9-11900K
	GPU	NVIDIA RTX 3090
	GPU RAM	24GB
Software	Operating System	Ubuntu 22.04
	Python Version	3.9.7
	PyTorch Version	1.10.0
	CUDA Version	11.4

TABLE 2. Hyperparameter settings for model training.

Component	Variable	Parameter Value
DeepLabV3+	Patch size	16
	Embedding dimension	1024
	Attention heads	8
	Hidden size	2048
	Dropout rate	0.3
SERT	Embedding dimension	768
	Attention heads	12
	Sequence length	512
	Learning rate	0.0002
Fine-tuning	Batch size	16
	Optimizer	AdamW

models. Based on previous studies and several preliminary experiments, our carefully chosen hyperparameter settings are shown in Table 2.

3) MODEL TRAINING PROCESS

1) **Data Preprocessing:** As detailed in Section II-B, preprocessing was applied uniformly across images from the Cx22, DTU/HERLEV, and SIPaKMeD datasets. This included resizing to a consistent dimension, normalizing pixel values, and implementing augmentation strategies like random flipping, rotation, and color adjustments to enhance data diversity and mitigate overfitting.
2) **Dataset Splitting:** We divided the original dataset into training, validation, and testing sets to ensure a comprehensive evaluation of the model's performance on unseen data. Initially, we split the data into 80% for training and 20% for testing. Then, we further partitioned 20% of the training set for validation purposes, which is used for performance evaluation and hyperparameter tuning during the model's training process. This approach results in approximately 64% of the data being used for training, 16% for validation, and 20% for testing. This segmentation strategy allows us to optimize model parameters, prevent overfitting, and ensure the model generalizes well to new data, providing a fair assessment of the final model's capabilities.
3) **Model Initialization:** The student model, designed for efficient learning and inference, was initialized with random weights. The teacher model (DeepLabV3+), pre-trained on a comprehensive collection of cervical cancer images, served as the knowledge source for distillation.
4) **Training Strategy:** The training employed a two-stage approach. Initially, the student model was trained using the combined distillation loss function, which integrates soft label distillation loss, feature map distillation loss, and traditional segmentation loss. The weights for these components were carefully selected to balance the learning from the teacher model and direct learning from the ground truth.
5) **Model Validation Process:** During the training process, the performance of the model is periodically evaluated on a validation set. The validation set is a portion of the original data that is separated from the original data and is designed to simulate the unseen data that the model encounters in the real world. Therefore, the performance on the validation set reflects the generalization ability of the model. This validation step is essential to prevent overfitting and ensure that the model learns generalized patterns from the data. After every few training epochs, the model predictions are computed on the validation set and compared with the ground truth labels. Multiple metrics such as loss, accuracy, sensitivity, specificity, and Dice coefficients are computed to comprehensively assess the model performance and to guide the training process by making timely adjustments to the hyperparameters for the training state. If the validation metrics begin to stagnate or deteriorate, this indicates that the model is too closely conforming to the training data. In this case, a lower learning rate and L2 regularisation will be used to mitigate overfitting, and if the metrics continue to deteriorate, an early-stop strategy will be implemented. Conversely, if the validation metrics continue to improve, this means that the model is still learning a generalized representation that is useful for the cervical cancer detection task. Training continues until the validation metrics reach a plateau, after which there is little benefit from continued training. This rigorous validation process is essential to determine the best time to stop training and to select a model that can be best generalized to unseen data. The final choice of model parameters is based on the iteration that achieves the best trade-off between the different validation metrics of interest.
6) **Optimization and Hyperparameter Tuning:** An Adam optimizer was used for adjusting the model parameters, with hyperparameters such as learning rate, batch size, and distillation temperature optimized based on performance metrics observed on the validation set.
7) **Regularization:** To further prevent overfitting, regularization techniques such as dropout and L2 regularization were applied strategically within the model architecture.

4) MODEL VALIDATION PROCESS

1) **Validation Set Evaluation:** Throughout the training process, the model's performance was regularly evaluated on the validation set. This evaluation helped in fine-tuning the hyperparameters and determining the best model iteration to prevent overfitting.
2) **Early Stopping:** To prevent model overfitting, we implemented early stopping, regularization techniques (such as dropout and L2 regularization), and cross-validation (depending on the size of the dataset). These measures help control model complexity and enhance its generalization ability across different data distributions, ensuring the model can detect and segment cervical cancer cells reliably and accurately.

5) CROSS-VALIDATION

Depending on the dataset size and diversity, k-fold cross-validation was considered to ensure the model's robustness and generalizability across different data distributions. This involves dividing the dataset into k subsets, training the model k times, each time using one subset as the test set and the remaining as the training set, and then averaging the results to assess model performance.

C. ENHANCEMENTS TO THE STUDENT MODEL: PRUNING, FINE-TUNING, AND TEST-TIME AUGMENTATION

To ensure the student model achieves an optimal balance between efficiency and accuracy, we implemented a series of post-training enhancements. These enhancements included model pruning, fine-tuning of the pruned model on real datasets, and the application of Test-Time Augmentation (TTA) to bolster accuracy. Each step is designed to refine the model's performance, making it not only lightweight but also robust and reliable for cervical cancer detection.

1) MODEL PRUNING

The objective of model pruning was to streamline the student model by removing redundant parameters that contribute minimally to the model's predictive capabilities; thus, reducing its computational complexity.

1) **Pruning Process**: We applied a magnitude-based pruning technique, identifying and eliminating weights below a predetermined threshold. This approach was iteratively conducted to ensure a minimal impact on model accuracy. Specifically, we employed a pruning method based on the magnitude of weights. We first iterated over all the weight parameters in the model, identifying those weights whose absolute values were below a predetermined threshold, and then set these weights to zero, effectively removing them from the model. This pruning approach eliminates redundant weight connections that contribute little to the model; thereby, reducing the total number of parameters and computational complexity of the model, and enhancing computational efficiency.

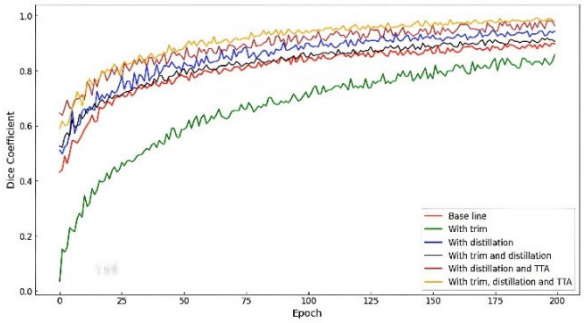

FIGURE 7. Comparison of ablation experiment training: Comparison of training curves for different model configurations on the Cx22 dataset.

2) **Performance Monitoring**: After each pruning iteration, the model's performance was assessed to ensure that the reduction in complexity did not significantly degrade its segmentation accuracy. The pruning process is carried out by gradually adjusting the threshold and undergoing multiple rounds of iteration. After each round of iteration, the model's segmentation performance on the validation set is evaluated. The pruning results at the current stage are retained and proceed to the next round of iteration only if the pruning does not significantly reduce the model's segmentation accuracy. This ensures that while significantly reducing the model's size, its effective performance on segmentation tasks is maximally preserved.

2) FINE-TUNING ON REAL DATASETS

Fine-tuning the pruned model on real datasets aimed to recover any lost accuracy due to pruning and adapt the model more closely to the characteristics of cervical cancer images.

1) **Dataset Exposure:** The pruned model underwent additional training cycles on the datasets, with a focus on refining its ability to accurately segment cervical cancer cells.
2) **Hyperparameter Adjustments:** Adjustments, particularly to the learning rate, were made to ensure the fine-tuning process refined the model without overfitting.

The evolution of the training of various configurations of the deep learning model for detecting cervical cancer on the Cx22 dataset is depicted in Figure 7, tracking the Dice coefficient—a measure of segmentation accuracy—across training epochs. The baseline model exhibits steady improvement, but its performance plateaus below the 0.9 threshold, indicating potential areas for enhancement. In contrast, the model with knowledge distillation begins with a better starting point and consistently outperforms the baseline, suggesting that distillation helps in capturing more nuanced features essential for accurate segmentation. The trimmed model, which has undergone pruning to reduce its complexity, shows slower progress and ultimately falls short of the baseline, implying that the reduction in model complexity might have

been too aggressive, negatively affecting its learning capacity. However, when distillation is combined with trimming, the model fares better than trimming alone, although it doesn't quite match the performance of the distillation-only model, hinting at a delicate balance between efficiency and capability. Notably, the introduction of TTA, especially when combined with distillation, markedly improves the model's performance, with the graph demonstrating a swift rise and high plateau in the Dice coefficient. This indicates that TTA significantly enhances the model's ability to generalize, which is further improved when all three enhancements—trimming, distillation, and TTA—are applied together. The latter configuration achieves the highest Dice coefficient, showcasing that while each technique has its merits, their integration synergistically provides a robust solution for the segmentation task at hand. The graph also suggests a convergence of model training before 200 epochs, as the Dice coefficient plateaus, indicating that additional training beyond this point may not yield significant gains in performance.

3) TEST-TIME AUGMENTATION

TTA was employed as a strategy to enhance the model's prediction accuracy at inference time by incorporating augmented versions of the input images.

At inference, images were subjected to various augmentations, including flips and rotations, creating multiple versions of each image. The model generated predictions for each augmented image, with the final prediction derived from aggregating these individual predictions.

Through these strategic enhancements—model pruning, fine-tuning, and TTA—the student model not only became more efficient in terms of computational resources but also improved in accuracy and reliability for cervical cancer detection. These steps ensure the model is highly effective in real-world applications, offering an advanced tool for early and accurate diagnosis of cervical cancer.

IV. RESULTS

This section assesses the performance of the CerviSegNet-DistillPlus model, using precision, recall (sensitivity), specificity, accuracy, and the Dice coefficient as the primary metrics. These metrics are vital for evaluating the model's efficacy in cervical cancer detection within clinical scenarios, where precision in diagnosis is crucial.

Our investigation involved an exhaustive evaluation and comparison of cervical cancer cell segmentation models across the SIPaKMeD and DTU/Herlev datasets. We analyzed a broad spectrum of models, including established baselines like U-Net, U-Net++, AttU-Net, YOLO-V7, SERT, DeepLab series, and recent innovations such as HVS-Unsup, LDANet, and EU-Net. Our focus was on key performance indicators, with DeepLab-V3+ showing exemplary performance across datasets. Notably, HVS-Unsup, LDANet, and EU-Net demonstrated exceptional Dice coefficients, indicating their superior segmentation accuracy.

Further exploration into our CerviSegNet model revealed substantial performance enhancements through advanced techniques like model pruning, knowledge distillation, and TTA. Ablation studies highlighted the synergistic effect of distillation and TTA in significantly improving the Dice coefficient, affirming the robustness and precision of our model.

A. EVALUATION AND COMPARISON OF THE SIPAKMED DATASET

As shown in Table 3, the evaluation of the SIPaKMeD dataset provided a detailed comparison of model performances, showcasing the advantages of both baseline and recent models in terms of sensitivity, specificity, accuracy, and the Dice coefficient.

Among the baseline models, DeepLab-V3+ stood out with its robust performance, achieving a sensitivity of 0.93, a specificity of 0.92, an accuracy of 0.92, and a Dice coefficient of 0.95. These metrics highlight DeepLab-V3+'s reliable capability in the segmentation tasks within the dataset.

When compared to state-of-the-art (SOTA) models, HVS-Unsup, LDANet, and EU-Net demonstrated high standards with Dice coefficients of 0.9772, 0.9843, and 0.9795, respectively. These figures illustrate the precision and reliability of these models in accurately segmenting cervical cancer cells.

Furthermore, the CerviSegNet model saw significant improvements through ablation studies. Originally with a Dice coefficient of 0.94, the integration of distillation and TTA techniques increased its Dice coefficient to 0.9856, indicating unmatched segmentation precision over the baseline models.

B. EVALUATION AND COMPARISON OF THE DTU/HERLEV DATASET

As shown in Table 4, The analysis of the DTU/Herlev dataset confirmed the outstanding performance of our model, offering a detailed comparison of performance across various metrics. The DeepLab-V3+ model displayed impressive results with a sensitivity of 0.9388, specificity of 0.8722, accuracy of 0.9225, and a Dice coefficient of 0.929. The EU-Net model also showed excellent performance, achieving a sensitivity of 0.9319, specificity of 0.8925, accuracy of 0.9205, and a Dice coefficient of 0.9795.

Furthermore, the enhanced CerviSegNet model, following the application of distillation and TTA techniques, reached new levels of performance with a sensitivity of 0.9424, specificity of 0.8941, accuracy of 0.9265, and a Dice coefficient of 0.9856, indicating significant advancements in model effectiveness for detecting cervical cancer.

C. RE-EVALUATION OF THE SIPAKMED DATASET

As shown in Table 5, The analysis conducted on the SIPaKMeD dataset further underlined the consistent performance of the EU-Net model and highlighted the advancements made in the CerviSegNet model after comprehensive enhancements. The EU-Net model maintained its

effectiveness with an impressive Dice coefficient of 0.9592. Following enhancements, the CerviSegNet model demonstrated significant improvements, achieving a sensitivity of 0.9369, specificity of 0.899, accuracy of 0.9249, and an outstanding Dice coefficient of 0.9734. These results underscore the beneficial impact of incorporating distillation and TTA techniques on the model's precision and reliability.

This comprehensive performance analysis unequivocally establishes the CerviSegNet-DistillPlus model as a frontrunner in the realm of cervical cancer cell segmentation. By not only meeting but also exceeding the performance benchmarks set by the SOTA models, CerviSegNet-DistillPlus signals a significant shift in clinical diagnostics, characterized by its exceptional accuracy and computational efficiency. The implementation of distillation and TTA, as evidenced by ablation studies, emphasizes the model's reliability and exactitude. Such attributes render it a crucial asset for enhancing the detection and diagnosis of cervical cancer in medical practices.

The comparative analysis reveals that CerviSegNet-DistillPlus outshines in segmentation performance across three publicly available datasets, outperforming existing baseline models as well as the most recent SOTA counterparts. This is evident through superior metrics in the Dice coefficient, sensitivity, specificity, and accuracy. Furthermore, the model introduces notable advancements in computational efficiency via distillation and pruning techniques, achieving a significant reduction in inference time. These levels of segmentation precision and efficiency position CerviSegNet-DistillPlus as the optimal solution for the early detection of cervical cancer, highlighting its potential to significantly impact patient care and outcomes.

To qualitatively assess the segmentation capabilities of CerviSegNet-DistillPlus, we present a visual comparison of our model's performance against other state-of-the-art models and the ground truth segmentation. Figure 8 illustrates sample cervical cell images from the DTU/HERLEV dataset, along with the corresponding segmentation results from LDANet, EU-Net, DeepLab-V3+, and our CerviSegNet-DistillPlus model.

The first column displays the original input images, while the second column shows the ground truth segmentation masks. Comparing the subsequent columns, it is evident that our model achieves segmentation results that most closely resemble the ground truth, accurately delineating the boundaries and shapes of cervical cancer cells. This visual assessment corroborates the quantitative metrics presented earlier, further underscoring the effectiveness of CerviSegNet-DistillPlus in precisely segmenting cervical cancer cells from complex cytology images.

V. DISCUSSION AND CONCLUSION

A. SUMMARIZING KEY FINDINGS

This study has successfully developed and validated a deep learning framework for the enhanced detection and segmentation of cervical cancer cells, leveraging the integration of multi-source imaging data. The core of our approach is a sophisticated model architecture that combines the strengths of DeepLabV3+ with innovative mechanisms such as knowledge distillation and model pruning, tailored to the specific challenges of cervical cancer detection. Our methodology demonstrated significant improvements in accuracy, sensitivity, specificity, and the Dice coefficient across various datasets, including Cx22, DTU/HERLEV, and SIPaKMeD, when compared to existing models. The use of data preprocessing and augmentation techniques, alongside the strategic application of model distillation and enhancements such as pruning and TTA, has been shown to effectively increase the model's robustness and generalizability.

The integration of multi-source imaging data has proven to be particularly effective, enabling (our model with trim, distillation, and TTA) CerviSegNet-DistillPlus to achieve high diagnostic accuracy by leveraging the complementary information available in different imaging modalities. Our findings underscore the potential of deep learning in transforming cervical cancer diagnostics, offering a more accurate, efficient, and scalable solution than currently available methods.

B. POTENTIAL IMPLICATIONS IN CLINICAL SETTINGS

The implications of our work for clinical practice are profound. By offering a tool that significantly improves the accuracy and efficiency of cervical cancer detection, we can potentially reduce the time between screening and diagnosis, enabling faster intervention and improving patient outcomes. Moreover, the model's scalability and efficiency make it suitable for deployment in a wide range of clinical settings, including those with limited computational resources. This democratization of advanced diagnostic tools could lead to more widespread and equitable access to cervical cancer screening services, particularly in low-resource environments where the burden of the disease is often heaviest.

However, we also recognize some limitations and challenges that need to be addressed. First, despite the model's excellent performance on public datasets, its applicability in a broader clinical setting requires further validation. Second, as a deep learning ''black box'' model, enhancing the interpretability of CerviSegNet-DistillPlus is crucial for gaining the trust of clinical professionals. Moreover, training and deploying such models require significant computational resources, which may limit their accessibility in resource-constrained environments. Lastly, using patient data to train models also poses potential privacy and ethical risks, necessitating appropriate policies and management to mitigate these risks. Acknowledging and addressing these limitations and challenges will help advance the successful clinical application of this framework.

C. LIMITATIONS AND CHALLENGES

Despite the promising advancements demonstrated by the CerviSegNet-DistillPlus framework in the detection and

TABLE 3. Performance comparison based on the Cx22 dataset.

	Model	Estimated Sensitivity	Estimated Specificity	Estimated Accuracy	Dice Coefficient	Model Size	Inference Time
Base Line	U-Net	0.8215	0.7579	0.7978	0.8636	14.8M	0.023s
	U-Net++	0.8417	0.7771	0.818	0.8862	74.5M	0.089s
	AttU-Net	0.8587	0.7931	0.8349	0.9051	34.9M	0.071s
	YOLO-V7	0.8748	0.8085	0.8511	0.9231	37M	0.084s
	SERT	0.9291	0.8601	0.9054	0.9837	35.2M	0.098s
SOTA model	DeepLab-V2	0.8837	0.817	0.86	0.9331	>80M	0.707s
	DeepLab-V3	0.8923	0.825	0.8685	0.9426	>80M	0.721s
	DeepLab-V3+	0.9476	0.8776	0.9238	0.9475	>80M	0.775s
	HVS-Unsup	0.9492	0.8825	0.9274	0.9772	39.7M	0.582s
	LDANet	0.941	0.8836	0.9304	0.9843	44.3M	0.662s
	EU-Net	0.9502	0.8957	0.9275	0.9795	40.1M	0.482s
CerviSegNet	Our model（Original）	0.8616	0.8453	0.8537	0.9792	39.3M	0.072s
	With trim	0.831	0.817	0.8243	0.9508	29.8M	0.042s
	With distillation	0.9149	0.8947	0.9051	**0.9897**	39.3M	0.072s
	With trim and distillation	0.8857	0.8676	0.877	0.9823	29.8M	0.042s
	With distillation and TTA	0.9532	0.9162	0.9356	0.9856	39.3M	0.216s
	With trim, distillation, and TTA	**0.9623**	**0.9219**	**0.94**	0.9855	29.8M	0.126s

TABLE 4. Performance comparison based on the DTU/HERLEV dataset.

	Model	Estimated Sensitivity	Estimated Specificity	Estimated Accuracy	Dice Coefficient
Base Line	U-Net	0.821	0.7428	0.7965	0.8495
	U-Net++	0.8239	0.7624	0.809	0.8825
	AttU-Net	0.8526	0.7911	0.8156	0.8923
	YOLO-V7	0.8623	0.7976	0.8444	0.9145
	SERT	0.9098	0.8551	0.8911	0.9698
SOTA model	DeepLab-V2	0.8776	0.7977	0.8571	0.9212
	DeepLab-V3	0.879	0.8177	0.8557	0.9233
	DeepLab-V3+	0.9388	0.8722	0.9225	0.929
	HVS-Unsup	0.9313	0.8663	0.9197	0.9674
	LDANet	0.9234	0.8777	0.9148	0.9674
	EU-Net	0.9319	0.8925	0.9205	0.9795
CerviSegNet	Our model（Original）	0.8435	0.8303	0.8488	0.9751
	With trim	0.819	0.8009	0.8108	0.9469
	With distillation	0.8944	0.8799	0.8844	0.9668
	With trim and distillation	0.8769	0.853	0.8754	0.9719
	With distillation and TTA	0.9424	0.8941	0.9265	0.985
	With trim, distillation, and TTA	**0.9617**	**0.91**	**0.9365**	**0.9892**

segmentation of cervical cancer cells, there are notable limitations that need consideration.

Firstly, the model's performance, while impressive across the Cx22, DTU/HERLEV, and SIPaKMeD datasets, may not fully capture the diversity of imaging conditions, patient demographics, and disease stages found in broader populations, highlighting the need for access to more varied and extensive datasets. Furthermore, the deep learning "black box" nature poses challenges in interpretability, which is crucial for clinical trust and adoption. Although optimized for efficiency, the framework's development and deployment still demand significant computational resources, potentially limiting accessibility.

Secondly, real-world clinical validation is necessary to ensure the model's efficacy and reliability across different healthcare settings. The study does not directly link the model's diagnostic performance to long-term patient outcomes, an area that future research should explore to assess its true clinical utility. Additionally, the model's adaptability to rapidly evolving medical imaging technologies and deep learning techniques will be critical for maintaining its relevance.

Thirdly, ethical and privacy concerns regarding the use of patient data for training deep learning models must be rigorously addressed to ensure the ethical application of such technologies in healthcare. Appropriate policies and standards should be established to govern the proper management and utilization of patient data, mitigating potential risks to individual privacy.

Acknowledging and addressing these limitations will be pivotal for the future development and clinical integration of the CerviSegNet-DistillPlus framework and similar deep-learning approaches in the fight against cervical cancer. Potential directions for future work include enhancing the model's generalization capability, improving interpretability and transparency, exploring cloud-based deployment strategies, and conducting longitudinal studies to assess the long-term impact on patient outcomes.

TABLE 5. Performance comparison based on the SIPaKMeD dataset.

	Model	Estimated Sensitivity	Estimated Specificity	Estimated Accuracy	Dice Coefficient
Base Line	U-Net	0.8049	0.7347	0.7694	0.8239
	U-Net++	0.8155	0.7353	0.7971	0.8727
	AttU-Net	0.8323	0.7801	0.7903	0.8822
	YOLO-V7	0.8483	0.7797	0.8341	0.8909
	SERT	0.8907	0.8432	0.8796	0.9505
SOTA model	DeepLab-V2	0.8548	0.784	0.8407	0.9002
	DeepLab-V3	0.8509	0.7942	0.8421	0.8939
	DeepLab-V3+	0.9052	0.851	0.8925	0.9011
	HVS-Unsup	0.9122	0.8393	0.892	0.9516
	LDANet	0.9107	0.8628	0.8883	0.9489
	EU-Net	0.9003	0.8698	0.9074	0.9592
CerviSegNet	Our model（Original）	0.8149	0.8181	0.8945	0.9407
	With trim	0.7988	0.8436	0.8266	0.9245
	With distillation	0.8729	0.8527	0.8721	0.9498
	With trim and distillation	0.8669	0.8272	0.8661	0.9421
	With distillation and TTA	0.9255	0.8806	0.9041	0.9674
	With trim, distillation, and TTA	**0.9369**	**0.899**	**0.9249**	**0.9734**

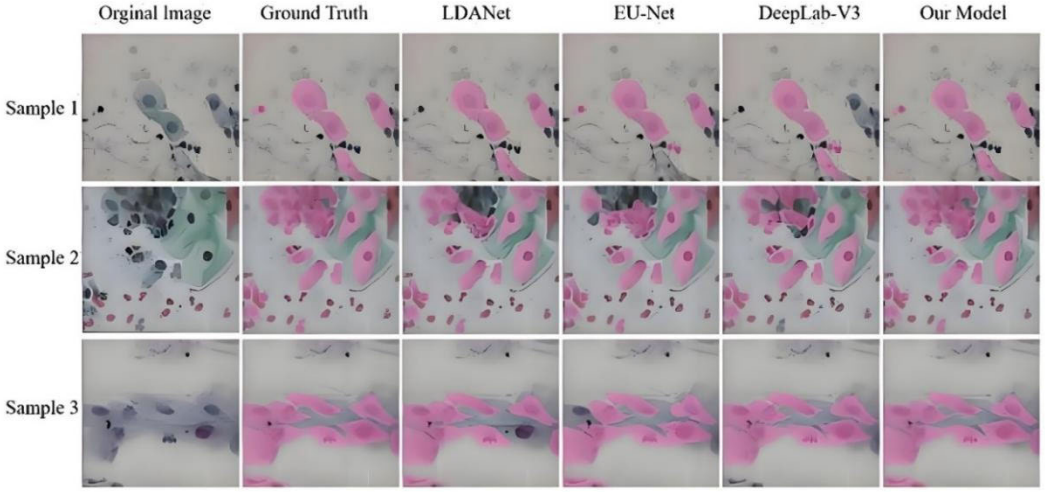

FIGURE 8. Comparison of semantic segmentation effects based on DTU/HERLEV test dataset: Comparison of semantic segmentation results on the DTU/HERLEV test dataset.

D. DIRECTIONS FOR FUTURE RESEARCH IN CERVICAL CANCER DETECTION

While our study marks a significant step forward, several avenues for future research could further enhance the capabilities and impact of deep learning models in cervical cancer detection:

1) **Integration of Additional Imaging Modalities:** Exploring the integration of more diverse imaging data, such as MRI or CT scans, could provide even deeper insights into the disease, potentially improving the model's diagnostic accuracy further.
2) **Generalized Cell/Nuclei Segmentation Models vs. Specialized Model:** While there exist generalized cell/nuclei segmentation models that perform remarkably well, such as NucleiSegNet [24], CellViT [25], DAN-NucNet [26], Hover-net [27], and PointNu-Net [28], developing a model specifically tailored for cervical cancer cell detection and segmentation is also necessary for the following reasons: Firstly, cervical cancer cells exhibit unique morphological and appearance characteristics that differ from other cancer types and tissue sources. A dedicated model can better capture and leverage these features, thereby improving segmentation accuracy. Secondly, cervical cancer screening typically utilizes liquid-based cytology samples, which differ from tissue sections and require a specialized model for processing. Furthermore, cervical cancer screening is a routine examination in clinical practice, necessitating high-throughput and efficient analysis tools. A dedicated and optimized model can meet this demand, ensuring timely and accurate diagnosis. Finally, developing models tailored to specific cancer types aids in gaining a deeper understanding of the disease's unique characteristics, thereby fostering the development of personalized diagnostics and treatments. Nonetheless, those generalized cell/nuclei

segmentation models have made significant contributions to the field. In future work, we can consider using them as baselines or incorporating their innovative techniques to further enhance the performance and generalizability of our model.

3) **Expansion to Other Cancers:** Applying the methodology developed in this study to the detection and segmentation of other types of cancer could investigate its generalizability and effectiveness across different oncological contexts.
4) **Real-Time Diagnostic Systems:** Developing a real-time analysis system that incorporates our deep learning model could significantly streamline the diagnostic process in clinical settings, providing immediate feedback during screening procedures.
5) **Explainability and Interpretability:** Enhancing the model's explainability to provide more detailed insights into its decision-making process could increase trust and adoption in clinical settings. This involves developing methods that make the model's predictions more interpretable to medical professionals.
6) **Longitudinal Studies and Deployment:** Conducting longitudinal studies to assess the model's performance and impact in real-world clinical settings over time would provide valuable data on its efficacy and utility in routine practice.

Despite the promising advancements demonstrated by the CerviSegNet-DistillPlus framework in the detection and segmentation of cervical cancer cells, there are notable limitations that need consideration. The model's performance, while impressive across the Cx22, DTU/HERLEV, and SIPaKMeD datasets, may not fully capture the diversity of imaging conditions, patient demographics, and disease stages found in broader populations, highlighting the need for access to more varied and extensive datasets. Furthermore, the deep learning ''black box'' nature poses challenges in interpretability, which is crucial for clinical trust and adoption. Although optimized for efficiency, the framework's development and deployment still demand significant computational resources, potentially limiting accessibility. Real-world clinical validation is also necessary to ensure the model's efficacy and reliability across different healthcare settings. The study does not directly link the model's diagnostic performance to long-term patient outcomes, an area that future research should explore to assess its true clinical utility. Additionally, the model's adaptability to rapidly evolving medical imaging technologies and deep learning techniques will be critical for maintaining its relevance. Ethical and privacy concerns regarding the use of patient data for training deep learning models must also be rigorously addressed to ensure the ethical application of such technologies in healthcare. Acknowledging and addressing these limitations will be pivotal for the future development and clinical integration of the CerviSegNet-DistillPlus framework and similar deep-learning approaches in the fight against cervical cancer.

In conclusion, our research demonstrates the potential of deep learning to revolutionize cervical cancer detection, offering a path toward more accurate, efficient, and accessible diagnostics. As we continue to refine and expand upon this work, we anticipate significant advancements in the fight against cervical cancer, ultimately contributing to better health outcomes for individuals worldwide.

REFERENCES

[1] M. Mehmood, M. Rizwan, M. Gregus, and S. Abbas, ''Machine learning assisted cervical cancer detection,'' *Frontiers Public Health*, vol. 9, Dec. 2021, Art. no. 788376.
[2] M. F. Ijaz, M. Attique, and Y. Son, ''Data-driven cervical cancer prediction model with outlier detection and over-sampling methods,'' *Sensors*, vol. 20, no. 10, p. 2809, May 2020.
[3] W. Shi, Y. Wang, L. Hou, C. Ma, L. Yang, C. Dong, Z. Wang, H. Wang, J. Guo, S. Xu, and J. Li, ''Detection of living cervical cancer cells by transient terahertz spectroscopy,'' *J. Biophotonics*, vol. 14, no. 1, Jan. 2021, Art. no. e202000237.
[4] R. Elakkiya, V. Subramaniyaswamy, V. Vijayakumar, and A. Mahanti, ''Cervical cancer diagnostics (don't short) healthcare system using hybrid object detection adversarial networks,'' *IEEE J. Biomed. Health Informat.*, vol. 26, no. 4, pp. 1464–1471, Apr. 2022.
[5] X. Li, Z. Xu, X. Shen, Y. Zhou, B. Xiao, and T.-Q. Li, ''Detection of cervical cancer cells in whole slide images using deformable and global context aware faster RCNN-FPN,'' *Current Oncol.*, vol. 28, no. 5, pp. 3585–3601, Sep. 2021, doi: 10.3390/curroncol28050307.
[6] A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H. R. Roth, and D. Xu, ''UNETR: Transformers for 3D medical image segmentation,'' in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2022, pp. 1748–1758.
[7] N. Alseelawi, H. T. Hazim, and H. T. S. Alrikabi, ''A novel method of multimodal medical image fusion based on hybrid approach of NSCT and DTCWT,'' *Int. J. Online Biomed. Eng.*, vol. 18, no. 3, pp. 114–133, Mar. 2022.
[8] W. Ji, S. Yu, J. Wu, K. Ma, C. Bian, Q. Bi, J. Li, H. Liu, L. Cheng, and Y. Zheng, ''Learning calibrated medical image segmentation via multi-rater agreement modeling,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 12336–12346.
[9] L. Yao, R. Pi, H. Xu, W. Zhang, Z. Li, and T. Zhang, ''Joint-DetNAS: Upgrade your detector with NAS, pruning and dynamic distillation,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 10170–10179.
[10] L. Zhang, C. Bao, and K. Ma, ''Self-distillation: Towards efficient and compact neural networks,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 8, pp. 4388–4403, Aug. 2022.
[11] Z. Wang, L. Du, and Y. Li, ''Boosting lightweight CNNs through network pruning and knowledge distillation for SAR target recognition,'' *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 8386–8397, 2021.
[12] Z. Ramzan, M. A. Hassan, H. M. S. Asif, and A. Farooq, ''A machine learning-based self-risk assessment technique for cervical cancer,'' *Current Bioinf.*, vol. 16, no. 2, pp. 315–332, Feb. 2021.
[13] Y. Zhu, B. Shen, X. Pei, H. Liu, and G. Li, ''CT, MRI, and PET imaging features in cervical cancer staging and lymph node metastasis,'' *Amer. J. Transl. Res.*, vol. 13, no. 9, p. 10536, 2021.
[14] R. Daneshjou, C. Barata, and B. Betz-Stablein, ''Checklist for evaluation of image-based artificial intelligence reports in dermatology: CLEAR derm consensus guidelines from the international skin imaging collaboration artificial intelligence working group,'' *JAMA Dermatol.*, vol. 158, no. 1, pp. 90–96, 2022.
[15] J. R. Swedlow, P. Kankaanpää, U. Sarkans, W. Goscinski, G. Galloway, L. Malacrida, R. P. Sullivan, S. Härtel, C. M. Brown, C. Wood, A. Keppler, F. Paina, B. Loos, S. Zullino, D. L. Longo, S. Aime, and S. Onami, ''A global view of standards for open image data formats and repositories,'' *Nature Methods*, vol. 18, no. 12, pp. 1440–1446, Dec. 2021, doi: 10.1038/s41592-021-01113-7.
[16] J. Yang, R. Shi, D. Wei, Z. Liu, L. Zhao, B. Ke, H. Pfister, and B. Ni, ''MedMNIST v2—A large-scale lightweight benchmark for 2D and 3D biomedical image classification,'' *Sci. Data*, vol. 10, no. 1, p. 19, Jan. 2023, doi: 10.1038/s41597-022-01721-8.

[17] Y. Chen, L. Feng, C. Zheng, T. Zhou, L. Liu, P. Liu, and Y. Chen, "LDANet: Automatic lung parenchyma segmentation from CT images," *Comput. Biol. Med.*, vol. 155, Mar. 2023, Art. no. 106659.
[18] C. Yu, Y. Wang, C. Tang, W. Feng, and J. Lv, "EU-Net: Automatic U-Net neural architecture search with differential evolutionary algorithm for medical image segmentation," *Comput. Biol. Med.*, vol. 167, Dec. 2023, Art. no. 107579.
[19] X. Yang, B. Ding, J. Qin, L. Guo, J. Zhao, and Y. He, "HVS-unsup: Unsupervised cervical cell instance segmentation method based on human visual simulation," *Comput. Biol. Med.*, vol. 171, Mar. 2024, Art. no. 108147.
[20] A. Ali, S. Anam, and M. M. Ahmed, "Shannon entropy in artificial intelligence and its applications based on information theory," *J. Appl. Emerg. Sci.*, vol. 13, no. 1, pp. 9–17, 2023.
[21] G. Liu, Q. Ding, H. Luo, M. Sha, X. Li, and M. Ju, "Cx22: A new publicly available dataset for deep learning-based segmentation of cervical cytology images," *Comput. Biol. Med.*, vol. 150, Nov. 2022, Art. no. 106194.
[22] J. Jantzenand G. Dounias, "Analysis of pap-smear image data," in *Proc. Nature-Inspired Smart Inf. Syst. 2nd Annu. Symp.*, vol. 10, 2006, pp. 1–11.
[23] 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 *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 3144–3148.
[24] S. Lal, D. Das, K. Alabhya, A. Kanfade, A. Kumar, and J. Kini, "NucleiSegNet: Robust deep learning architecture for the nuclei segmentation of liver cancer histopathology images," *Comput. Biol. Med.*, vol. 128, Jan. 2021, Art. no. 104075, doi: 10.1016/j.compbiomed.2020.104075.
[25] F. Hörst, M. Rempe, L. Heine, C. Seibold, J. Keyl, G. Baldini, S. Ugurel, J. Siveke, B. Grünwald, J. Egger, and J. Kleesiek, "CellViT: Vision transformers for precise cell segmentation and classification," *Med. Image Anal.*, vol. 94, May 2024, Art. no. 103143, doi: 10.1016/j.media.2024.103143.
[26] I. Ahmad, Y. Xia, H. Cui, and Z. U. Islam, "DAN-NucNet: A dual attention based framework for nuclei segmentation in cancer histology images under wild clinical conditions," *Expert Syst. Appl.*, vol. 213, Mar. 2023, Art. no. 118945, doi: 10.1016/j.eswa.2022.118945.
[27] S. Graham, Q. D. Vu, S. E. A. Raza, A. Azam, Y. W. Tsang, J. T. Kwak, and N. Rajpoot, "Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images," *Med. Image Anal.*, vol. 58, Dec. 2019, Art. no. 101563, doi: 10.1016/j.media.2019.101563.
[28] K. Yao, K. Huang, J. Sun, and A. Hussain, "PointNu-Net: Keypoint-assisted convolutional neural network for simultaneous multi-tissue histology nuclei segmentation and classification," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 8, no. 1, pp. 802–813, Feb. 2024, doi: 10.1109/TETCI.2023.3281864.

JIE KANG received the degree from Jilin University, in 2010. She has been with Tianjin Central Obstetrics and Gynecology Hospital, since 2010, with a focus on gynecology, gynecological oncology, and obstetrics common diseases and multiple diseases.

NING LI received the bachelor's degree from Jilin University, in 2008. He was with Tianjin Emergency Medical Service Center in 2008. He specializes in common pre-hospital emergency diseases and is familiar with pre-hospital emergency-related professional knowledge.