scientific reports

Check for updates

OPEN

CytoNet: an efficient dual attention based automatic prediction of cancer sub types in cytology studies

Naveed Ilyas[1], Farhat Naseer[2], Anwar Khan[3], Aamir Raja[4], Yong-Moon Lee[5], Jae Hyun Park[6] & Boreom Lee[7]✉

Computer-assisted diagnosis (CAD) plays a key role in cancer diagnosis or screening. Whereas, current CAD performs poorly on whole slide image (WSI) analysis, and thus fails to generalize well. This research aims to develop an automatic classification system to distinguish between different types of carcinomas. Obtaining rich deep features in multi-class classification while achieving high accuracy is still a challenging problem. The detection and classification of cancerous cells in WSI are quite challenging due to the misclassification of normal lumps and cancerous cells. This is due to cluttering, occlusion, and irregular cell distribution. Researchers in the past mostly obtained the hand-crafted features while neglecting the above-mentioned challenges which led to a reduction of the classification accuracy. To mitigate this problem we proposed an efficient dual attention-based network (CytoNet). The proposed network is composed of two main modules (i) Efficient-Net and (ii) Dual Attention Module (DAM). Efficient-Net is capable of obtaining higher accuracy and enhancing efficiency as compared to existing Convolutional Neural Networks (CNNs). It is also useful to obtain the most generic features as it has been trained on ImageNet. Whereas DAM is very robust in obtaining attention and targeted features while negating the background. In this way, the combination of an efficient and attention module is useful to obtain the robust, and intrinsic features to obtain comparable performance. Further, we evaluated the proposed network on two well-known datasets (i) Our generated thyroid dataset (ii) Mendeley Cervical dataset (Hussain in Data Brief, 2019) with enhanced performance compared to their counterparts. CytoNet demonstrated a 99% accuracy rate on the thyroid dataset in comparison to its counterpart. The precision, recall, and F1-score values achieved on the Mendeley Cervical dataset are 0.992, 0.985, and 0.977, respectively. The code implementation is available on GitHub. https://github.com/naveedilyas/CytoNet-An-Efficient-Dual-Attention-based-Automatic-Prediction-of-Cancer-Sub-types-in-Cytol.

The application of artificial intelligence (AI)[2–4] gets challenged when dealing with heterogeneous diseases that may lead to death. Cancer is the second most deadly disease worldwide, especially in the USA, ranking second in deaths in the United States. The World Health Organization is expecting a total number of 29.4 million new cancer cases by 2040. Early diagnosis of cancer leads to curing patients to some extent. To diagnose cancer whole-slide images (WSI) are commonly observed by pathologists through the microscope to differentiate between malignant and normal tissues. The manual diagnostic process used by pathologists is prone to different errors based on different pathologists and specimen types, type-I and type-II errors occur in 6% and 33% of cancer cases, respectively. CAD of WSI is a complicated process due to the nature of the cell's biological morphology, thus the traditional machine learning process may fail to provide a generalized solution even coupled with handcrafted features.

[1]Department of Bioengineering, University of California Riverside, Riverside, CA 92521, USA. [2]School of Mechanical and Manufacturing Engineering (SMME), National University of Science and Technology (NUST), Islamabad, Pakistan. [3]VIB-KU Leuven Center for Cancer Biology, Katholieke Universiteit Leuven, UZ Leuven, Leuven, Belgium. [4]Department of Physics, Khalifa University of Science and Technology, Abu Dhabi, UAE. [5]Department of Pathology, College of Medicine, Dankook University, Cheonan, South Korea. [6]Department of Surgery, Wonju Severance Christian Hospital, Wonju College of Medicine, Yonsei University, Wonju, South Korea. [7]Department of Biomedical Science and Engineering, Gwangju Institute of Science and Technology, Gwangju, South Korea. ✉email: leebr@gist.ac.kr

Image analysis has lot of applications starting from terrestrial to underwater imaging using sensor networks[5–7]. The recent advancement in Convolutional Neural Networks (CNNs)[8–11] is useful in a variety of applications ranging from computer vision, language processing and medical image analysis. It is believed that CNN is capable of more accurately diagnosing cancer as compared to traditional methods. Compared to traditional machine learning approaches, CNN-based methods have a very data-hungry nature which leads to labeling the large number of WSI. Annotating the WSI is a very time-consuming and difficult task due to many reasons such as human bias, technical issues and the availability of professional pathologists. To ease the annotating process, pathologists annotate the abnormal regions in WSI while other regions are automatically considered benign.

For the classification of different types of tumors, a number of machine-learning techniques have been proposed. According to our knowledge, the first deep semantic mobile application was built [12] by using four different types of traditional machine learning algorithms such as Random Forest (RM), linear Support Vector Machine (SVM), k-Nearest Neighbour (KNN), and Convolutional Neural Networks (CNNs). Whereas, Deep Neural Networks (DNNs) such as VGG-16 and Inception-v3 [13] were used by authors in 2019. Similarly, authors in [14] used Inception-v2 and VGG-19 for multi-class classification of thyroid carcinoma using histology images. Authors in [15] used two cascaded CNNs to classify thyroid carcinoma in WSI. Both networks contributed to a cascaded network based on VGG-11. Whereas, authors in [16] used two CNNs to predict the malignancy. Tao et al. [17] built a system for follicular segmentation of thyroid cytopathology by using WSI. They applied the ResNet101 over the hybrid segmentation algorithm to enhance the accuracy.

In recent years, significant strides have been made in the field of computer vision, particularly in the context of medical image analysis. Notable among these advancements are the development and application of deep learning architectures such as VGGNet, ResNet, and attention-based models. These architectures, as evidenced by studies such as Dov et al. (2019)[15], Tao et al. (2019)[17], and Oksuz et al. (2024)[18], have demonstrated remarkable efficacy across various tasks within the domain of computer vision.

However, despite their successes, it has become increasingly apparent that these architectures may not be optimally tailored to address the inherent complexities of Whole Slide Image (WSI) analysis, particularly in the context of cancer diagnosis. The primary limitation lies in their design, which may inadvertently overlook critical spatial and contextual information vital for the accurate classification of cancerous cells. In the intricate landscape of WSI, where subtle morphological variations and spatial relationships play a pivotal role in diagnosis, the simplistic architectures of VGGNet, ResNet, and even attention-based models may fall short in capturing the nuanced features essential for precise classification. This deficiency poses a significant challenge to Computer-Aided Diagnosis (CAD) systems, which rely heavily on accurate feature extraction and classification for timely and effective diagnosis and treatment planning. Failure to adequately discern between different types of carcinomas due to limitations in feature representation can impede the overall efficacy of CAD systems, thereby potentially delaying crucial interventions and compromising patient outcomes.

Building upon the insights gleaned from the aforementioned literature and empirical observations, we propose CytoNet: an innovative framework designed specifically to address the shortcomings of existing architectures in the context of papillary carcinoma prediction using fine needle aspiration cytology. CytoNet represents a paradigm shift in automated cancer prediction, leveraging a synergistic combination of cutting-edge techniques to enhance feature extraction and classification accuracy. At the core of CytoNet lies a novel architecture comprising two key components: Efficient-Net (B0) and the Channel Attention Module (CAM). The choice of Efficient-Net (B0) stems from its well-established reputation for achieving exceptional accuracy through a principled approach to model scaling, encompassing adjustments in width, depth, and resolution. By leveraging the power of compound scaling, Efficient-Net (B0) ensures that CytoNet can effectively capture both low-level and high-level features essential for accurate cancer subtype prediction. Complementing the prowess of Efficient-Net (B0) is the Channel Attention Module (CAM), a sophisticated mechanism tailored to extract target-oriented features while attenuating background noise. CAM's ability to selectively focus on salient regions within the image enhances the visibility of crucial diagnostic features, thereby facilitating more accurate classification. By incorporating two CAM modules in parallel, CytoNet capitalizes on synergistic feature extraction, culminating in a robust representation of intrinsic features critical for cancer subtype prediction.

In summary, the main contributions of our research are as follows:

- We have designed an efficient net-based deep learning architecture for accurately identifying different carcinoma types.
- The efficient baseline model is very effective in obtaining rich features with reduced computational cost.
- We have evaluated the proposed method on two datasets (thyroid, cervical) which gave a comparable performance when compared with state-of-the-art methods.The study focuses on the classification of thyroid and cervical data. The research begins by explaining the importance and challenges associated with accurately categorizing thyroid and cervical datasets. Subsequently, relevant literature and methodologies related to the classification of thyroid and cervical data are reviewed. The data preprocessing section outlines the steps taken to prepare the datasets for analysis. The proposed classification model is then presented, detailing its architecture and training procedures. Experimental results, including performance metrics and comparisons with existing methods, are subsequently discussed. The implications of the findings and potential limitations are also addressed. Finally, the paper concludes by summarizing key findings and suggesting directions for future research in thyroid and cervical classification.

Related work

With the rapid growth of CNN-based techniques in classification, recognition, and especially segmentation tasks, CNN-based methods are employed for the purpose of medical image analysis. To address challenges such

as structural variation and a small number of training samples, a number of researchers have contributed to enhancing segmentation accuracy. Further, the table. 1 depicts the key features and summary of each technique.

A detailed review is wrote by authors in[36] wrote a detailed review on the automatic classification and segmentation of cervical cancer. They have reviewed the state-of-the-art techniques published on well-known venues. By reviewing the different techniques, some of the most robust and efficient techniques are highlighted. Also, the drawbacks and deficiencies of each technique are explored to motivate the researcher to dive deep into a specific research room. A comparison is drawn among ResNet-50, random forest, and ensemble learning to check the diagnostic efficiency for papillary thyroid carcinoma (PTC) in[37]. They analyzed 559 patients' datasets by collecting thyroid pathological images. Whereas, authors in[14] presented a deep convolutional neural network (DCNN), more specifically VGG-19[38] to classify normal tissue, adenoma, nodular goitre, papillary thyroid carcinoma (PTC), follicular thyroid carcinoma (FTC), medullary thyroid carcinoma (MTC) and anaplastic thyroid carcinoma (ATC). An intelligent simple Deep Neural Network (DNN) to classify the cervical cancerous samples in whole slide image is proposed in[39]. The architecture is composed of four types of layers: Convolutional, activation, pooling, and fully connected layers. The proposed model achieved 93.3% accuracy with a reasonable improvement compared to its counterpart. Two main categories of DNN are self-supervised and supervised learning. Each has its own advantages and disadvantages. For example, at the cost of high computational resources, unlabeled data is used in the self-supervised technique. Whereas, prior data information is acquired at the cost of expert annotation time. Thus, authors in[19] proposed a hybrid technique to take advantage of both self-supervised and supervised techniques to increase classification performance. Authors in[20] explores cervical cancer classification using datasets from the Intel and MobileODT Cervical Cancer Screening competition on Kaggle. It investigates training standard DNNs with weights from pre-training on the ImageNet dataset for multi-class classification, suggesting potential with smaller DNN versions. Additionally, it examines the use of the EfficientNet-B0 model for optimizing image size in multi-label classification. The approach offers a useful strategy for image size optimization in various scenarios. Authors in[21] proposed a transfer learning-based approach to classify thyroid cancer. They have applied the most popular CNNs (DenseNet121 and NasNetLarge) for the classification of cancer by using ultrasound images. A simple Conv-Net method to categorize thyroid nodule malignancy by using ultrasound images is proposed in[22]. The main purpose is to differentiate the regular and irregular lumps. They have also performed pre-processing, segmentation, feature extraction, and data balancing. Whereas, classification was performed through deep neural networks. Authors in[23] proposed a multi-modal feature fusion approach to classify benign and malignant categories. The fusion of different

Reference	Key features	Summary
Lubran[19]	Hybrid technique combining self-supervised and supervised learning	Presented a hybrid technique combining self-supervised and supervised learning to enhance classification performance.
Tomko[20]	Training standard DNNs and EfficientNet-B0 model	Investigated cervical cancer classification using datasets from Intel and MobileODT Cervical Cancer Screening competition, exploring training standard DNNs and EfficientNet-B0 model.
Nugroho[21]	Transfer learning-based approach using DenseNet121 and NasNetLarge	Proposed a transfer learning-based approach using DenseNet121 and NasNetLarge for classifying thyroid cancer from ultrasound images.
Nugroho[22]	Conv-Net method for categorizing thyroid nodule malignancy	Presented a Conv-Net method for categorizing thyroid nodule malignancy from ultrasound images, incorporating pre-processing, segmentation, and feature extraction.
Avola[23]	Multi-modal feature fusion approach	Introduced a multi-modal feature fusion approach for classifying benign and malignant thyroid nodules to improve accuracy.
Kanavati[24]	Classification of cancerous and non-cancerous Whole Slide Images (WSI) using EfficientNet-B0	Investigated the classification of cancerous and non-cancerous Whole Slide Images (WSI) using EfficientNet-B0.
Sornapudi[25]	Image analysis toolbox based on deep learning models	Developed an image analysis toolbox based on deep learning models for early diagnosis of cervical cancer, including detection, segmentation, analysis, and classification tasks.
Arifianto[26]	Pre-trained SqueezeNet for multi-class classification	Utilized pre-trained SqueezeNet for multi-class classification of cervical cancer using whole slide images.
Habtemariam[27]	Dual network using MobileNetv2-YOLOv3 and EfficientNetb0	Presented a dual network using MobileNetv2-YOLOv3 for detection and EfficientNetb0 for classification of cervical cancer.
Attallah[28]	Novel computer-aided diagnostic (CAD) system	Proposed a novel CAD system for cervical cancer detection, extracting features from multiple domains for enhanced performance.
Attallah[29]	CerCan-Net with transfer learning	Introduced CerCan-Net, a CNN-based approach with transfer learning for automatic cervical cancer diagnosis and feature selection.
Shinde[30]	Deep learning-based hybrid methodology for pap smear cytology image classification	Introduced DeepCyto, a deep learning-based methodology for classifying pap smear cytology images, addressing the challenge of analyzing numerous slides.
Liao[31]	Automatic system for recognizing cervical cancerous images using whole slide images	Presented an automatic system for recognizing cervical cancerous images using whole slide images, incorporating sliding detector and lesion analyzer modules.
Li[32]	Hybrid technique combining self-supervised learning with multiple instances learning	Proposed a hybrid technique combining self-supervised learning with multiple instances learning for cervical cancer classification using whole slide images.
Buddhavarapu[33]	Automatic thyroid image classification using deep neural networks and transfer learning	Presented an automatic thyroid image classification using deep neural networks and transfer learning with various CNN architectures.
Liu[34]	Deep learning framework for cervical cell classification	Developed a deep learning framework for cervical cell classification in cytopathology images, combining CNN and Visual Transformer modules.
Rahman[35]	DeepCervix with hybrid deep feature fusion (HDFF)	Introduced DeepCervix, a deep learning-based method utilizing hybrid deep feature fusion (HDFF) for accurate cervical cell classification.

Table 1. Summary of related work.

image representations is useful to better classify thyroid nodules by enhancing accuracy. Authors in[24] proposed a ConvNet to investigate the classification of cancerous and non-cancerous WSI. They specifically used the EfficientNet-B0 with an input size of 1024×1024. They evaluated the proposed approach on three test datasets with considerably good results. Precancerous stage is somehow very difficult to detect. Accurate detection of the precancerous stage is very helpful in determining the patient's risk for developing the disease. The samples are taken through biopsy and examined by the radiologist using a microscope. However, manual examination is prone to error, time taking, and intra-inter-observer variability. Thus, authors in[25] presented a method to detect the precancerous state of the uterine cervix. More precisely, they have developed an image analysis toolbox for early diagnosis of cervical cancer. The toolbox is based on deep learning models with different tasks i.e, (i) Detect the cancerous region (ii) Segment the detected portion, (iii) Analyze the region and (iv) Classification with localization of the detected region. A pre-trained SqueezeNet architecture for multi-class classification of cervical cancer using whole slide images by obtaining a comparable performance[26]. Authors in[27] presented a dual network for the detection and classification of cervical cancer. The lightweight MobileNetv2-YOLOv3 was used for the detection of a region of interest (ROI), whereas pre-trained EffecientNetb0 was used for performing the classification. Authors in[28] proposed a novel computer-aided diagnostic (CAD) system for cervical cancer detection. Unlike existing methods, their approach extracts features from multiple domains, eliminating the need for pre-segmentation and enhancing performance. By employing compact deep learning models and extracting statistical and textural descriptors from various domains, the proposed CAD achieves exceptional accuracy, reaching 100% with only 35 principal components. Comparative analysis demonstrates the superior performance of the proposed CAD, showcasing its potential for enhancing cervical cancer diagnosis. Authors in[29] introduced CerCan-Net, a new approach for automatic cervical cancer diagnosis. Utilizing three lightweight CNNs and transfer learning, CerCan-Net extracts deep features to enhance performance. Through feature selection, it differentiates subgroups of cervical cancer. The study positions CerCan-Net as a promising tool for assisting cytopathologists and addressing routine diagnostic challenges.

To diagnose cervical cancer, a pap smear test is one of the most preferred ones. It is time-consuming as experts analyze 100 to 1000 slides. Due to this challenge, authors in[30] presented a deep learning-based hybrid methodology named as DeepCyto for the classification of pap smear cytology images. An automatic system to recognize cervical cancerous images using whole slide images is presented in[31]. They have used a sliding detector and lesion analyzer module to increase the overall performance. A hybrid technique by combining self-supervised with multiple instances learning to classify cervical cancerous using whole slide images is presented in[32]. They have used the pre-trained testing based self-supervised model for classification purposed by gaining a reasonable performance. Whereas, authors in[33] presented an automatic thyroid image classification by using deep neural networks using transfer learning and pre-trained CNNs. They used InceptionNet, ResNet, VGGNet and DenseNet for feature extraction and transfer learning to classify the histopathology images. Authors in[34] proposed a deep learning framework for cervical cell classification in cytopathology images. It combines CNN and Visual Transformer modules for feature extraction and a Multilayer Perceptron for classification. Whereas authors in[35] presented a DeepCervix, a novel deep learning-based method for accurate cervical cell classification. Unlike previous approaches, DeepCervix employs a hybrid deep feature fusion (HDFF) technique to enhance classification accuracy.

Dataset preparation and pre-processing

This research is conducted with the approval of the Institutional Review Board (IRB) of Wonju Severance Christian Hospital, Wonju College of Medicine (IRB Approval No. CR320109), and Yonsei University. The cytological images were obtained from patients who had thyroid carcinomas. The images were obtained using fine needle aspiration cytology (FNAC), and thyroidectomy from January 1st, 2013 to December 31st, 2019, respectively. Before performing FNAC, written informed consent was obtained from all the participants. All methods and procedures while obtaining FNAC were performed in accordance with the relevant guidelines and regulations. The FNAC is performed with 22-gauge needles by an experienced sonographer with eight years of experience in thyroid ultrasound examination. Also, he is a member of the Korean Surgical Ultrasound Society. Thin-layer liquid-based cytology (LBC) preparations are far better as compared to traditional preparations by considering cell preservation, mono-layer cell preparation, and background clarity. Therefore, most of the FNAC samples were transferred to a 10 ml syringe and prepared with a natural sedimentation-type thin layer LBC system using a BD SurePath liquid-based Pap Test (Beckton Dickinson, Durham, NC, USA). Whereas, a digital still camera with specifications (DP27, Olympus, Tokyo, Japan), 40× objective lens attached to BX45 Olympus microscope is used to take pictures for the LBC smears. The images are annotated and saved in JPEG format by two experienced cytopathologists having five to ten years of experience in thyroid cytopathologic examination. Our thyroid dataset contains 367 hematoxylin-eosin (H & E)-stained images. The dataset is composed of 222 patients with papillary thyroid carcinoma (PTC), and 145 patients with non-PTC. The obtained slides are digitized at 400× magnification. The PTC images with following features cellularity, papillary fronds with anatomical edges, nuclear crowding and overlapping, chewing gum colloid, enlarged oval nuclei with longitudinal intranuclear grooves, and cellular swirls[40]. And these images are categorized in category VI by following the Bethesda system for reporting thyroid cytopathology[41]. The images with benign nodules fit the category II description in accordance with the Bethesda system. And it consists of a cellular specimen comprises of varying proportions of colloid and benign follicular cells arranged as macrofollicles and macrofollicle fragments[41]. Subjects in category II who underwent thyroidectomy were pathologic diagnoses of benign thyroid lesions. Detailed information about the original dataset with the number of balanced fragments for training, validation, and testing is shown in Table 2.

Cytological type	OI	OGF	RSF	TF	VF	Test-F
PTC (malignant)	222	866	700	495	140	70
Non-PTC (benign)	145	713	700	495	140	70
All types	367	1579	1400	980	280	140

Table 2. Numbers of original FNAC images in each class, and numbers of automatically extracted fragments in each randomly selected balanced sub-sample. Original image (OI), Originally generated fragments (OGF), Randomly selected fragments (RSF), Training fragments (TF), Validation fragments (VF), and Test fragments (Test-F)

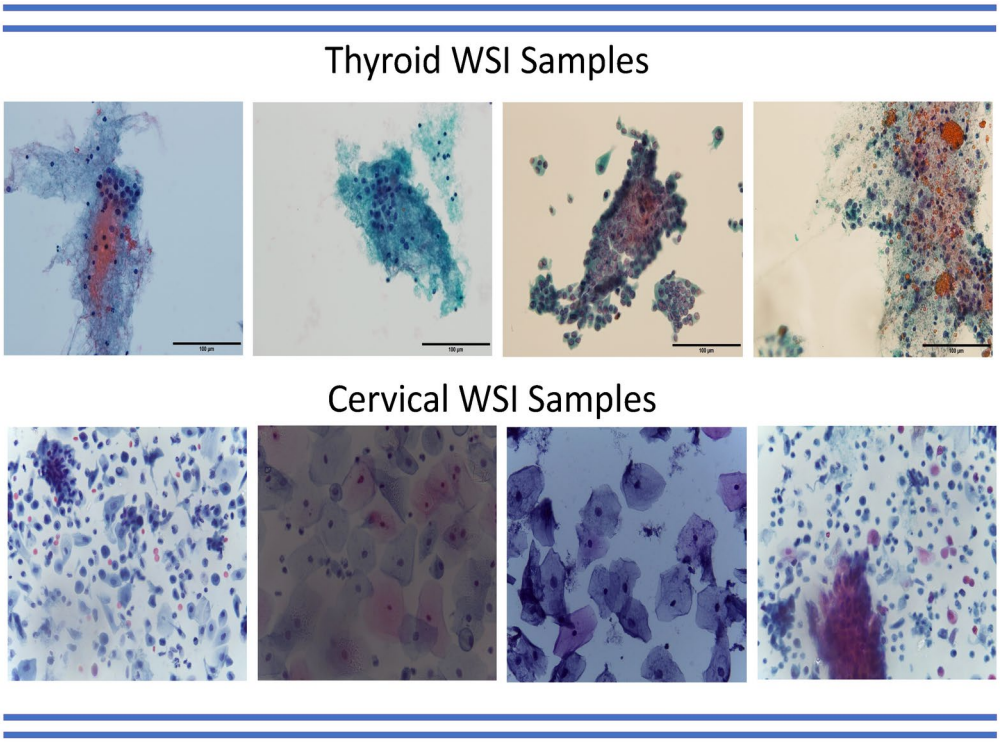

Figure 3. Visualization of samples from two different datasets: (i) thyroid dataset (own), and (ii) Cervical dataset (Mendeley).

Pre-processing and automatic generation of fragment images

Different microscopic systems are used to obtain cytological images which lead to varying background and colour fluctuation. Four random samples were selected from thyroid and cervical cancer datasets with different backgrounds and colours as shown in Fig. 3. To normalize the dataset, different normalization techniques were adopted. We have therefore used stain colour normalization[42]. A simple statistical analysis is performed to impose one image (source) on another reference image. Afterward, stain normalization on the source image is obtained by applying its characteristics to the reference image. Original cytology images of 4800 × 3600 pixels are divided into a number of patches (fragments), each containing non-overlapping clusters of tissues or region of interest (ROI). By considering the opinion of an experienced cytopathologist CAD contains at least more than five cells to ensure the reliability of clinical diagnosis. In this way, not only does the number of training examples increase but also it also performs the practical-oriented diagnosis in the current machine/deep learning-based ROI system. We thus fragmented the original image into a size of 200 × 200 pixels to maintain enough informative tissue structure. The re-annotation of the fragments is performed to maintain data quality, thus obtaining more accurate and reliable labels for further analysis. The dissimilar fragments with the original FNAC slide are discarded to further ensure the quality of the thyroid dataset.

Proposed model

The architecture of the proposed approach is shown in Fig. 1. WSI images were obtained after the experiment. The WSI is divided into small patches after applying normalization. The normalized patches are fed into the proposed model. Firstly, the normalized image patches are passed through a pre-trained Efficient Net (B0). Efficient Net (B0) is very useful to obtain general to detailed features. Further, to obtain the most relevant features while negating the irrelevant part, we placed two channel attention modules (CAM) in parallel as shown in Fig.

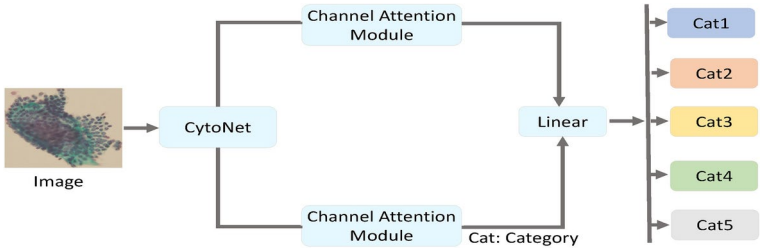

Figure 1. An overview of CytoNet: the proposed network is composed of two main modules; (1) Efficient Net, (2) Dual attention module (DAM).

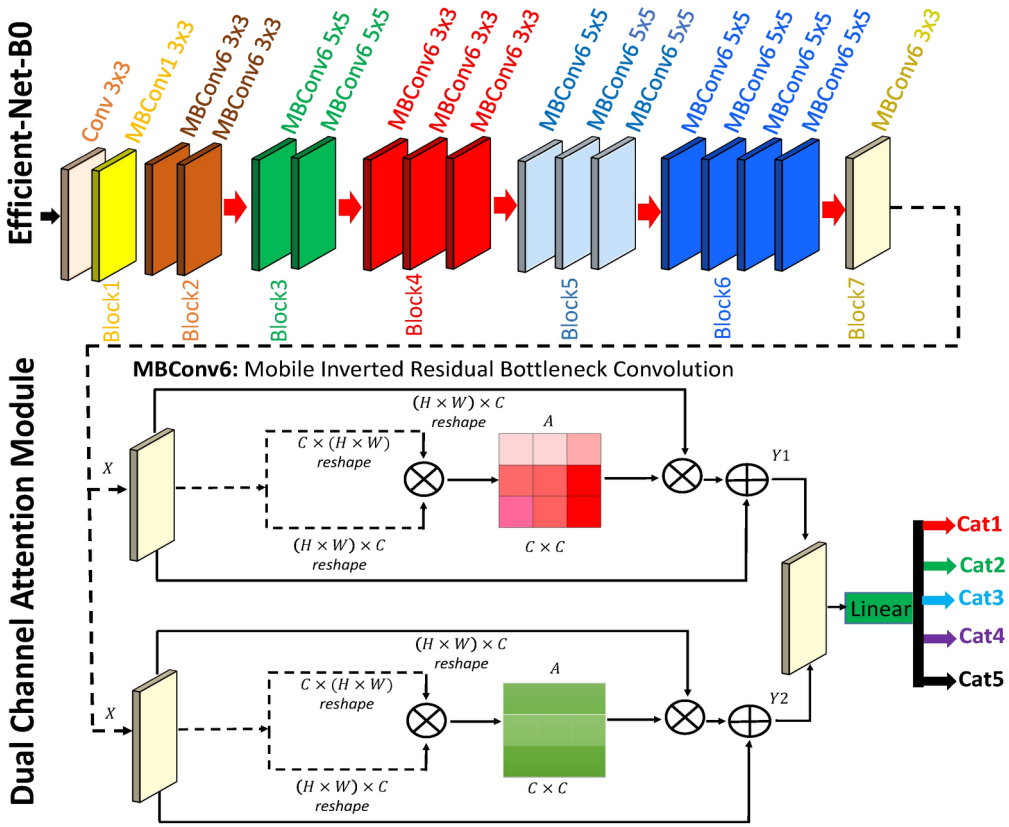

Figure 2. CytoNet: an efficient dual-attention based automatic prediction of papillary carcinoma using fine needle aspiration cytology. The proposed network composed of two main modules: (1) EfficientNet-B0 (top), and channel attention module (bottom). The input image first enters into EfficientNet-B0 and is forwarded to the dual channel attention module.

1. After CAM, we used the dropout layer to proceed with the linear layer to classify the different categories of cancer.

CytoNet: an efficient dual-attention based automatic prediction of papillary carcinoma using fine needle aspiration cytology

EfficientNet-B0

Convolutional neural networks (CNNs) are used to obtain detailed features with a common concept. This concept is described in terms of accuracy which is supposed to be directly proportional to the available resources. It can be justified with a number of examples, for example, ResNet can be scaled up to ResNet-18 and 200, respectively (increasing the number of layers).

The traditional method of increasing the accuracy is by increasing the depth and width of the network. These methods have proven to be successful. However, they require tedious manual tuning. Instead, a novel method is used to obtain better accuracy and efficiency. Compared to conventional methods that arbitrarily scale network dimensions like depth, width, and resolution, a uniform scaling of the dimension with a fixed set of coefficients is more effective. These types of networks have their own family of models. We have used the

pre-trained EfficientNet-B0. The core concept of EfficientNet-B0 is compound scaling. It is defined as scaling individual dimensions by balancing the network width, depth and resolution against available resources to improve performance. The EfficientNet-B0 comprises seven blocks starting from block 1 and block 7. Block 1 is borrowed from MobileNet inverted bottleneck convolution with a convolutional layer of size 3 × 3. Whereas, the same mobile inverted bottleneck convolution is used in block 2 with two convolutional layers with a kernel size of 3 × 3. Similarly, block 3, 4, 5, 6 and 7 are also motivated by mobile inverted bottleneck convolution with convolutional layers of 2, 3, 3, 4, and 1 respectively. However, different kernel sizes are used in each of the blocks as shown in Fig. 2. Mobile Inverted Residual Bottleneck Convolution (MBConv) The MBConv block consists of three main components:

- Depth-wise Convolution: This step applies separate convolutions to each input channel independently. It helps reduce the computational cost by performing lightweight filtering on individual channels.
- Point-wise Convolution: Following the depth-wise convolution, a 1 × 1 point-wise convolution is applied to linearly combine the output channels from the depth-wise convolution. This step helps capture cross-channel correlations and enables the model to learn more complex representations.
- Bottleneck Structure: The MBConv block employs a bottleneck design, which reduces the dimensionality of the input feature maps using a 1 × 1 point-wise convolution with fewer output channels. This dimension reduction helps further reduce the computational burden while maintaining the representational capacity of the model.By stacking multiple MBConv blocks with increasing complexity and width, Efficient-Net models achieve impressive performance while being computationally efficient. The MBConv block plays a critical role in enabling Efficient-Net to efficiently scale up to different model sizes and achieve state-of-the-art results in cervical and thyroid classification tasks.

Channel attention module (CAM)

A computationally economical Efficient-Net-B0 plays a vital role in enhancing classification accuracy. However, the irrelevant regions in image patches need to be ignored to enhance overall performance. To highlight targeted areas, we used dual CAM. There are two types of class-specific responses in the cancer classification domain: i) cancerous cell, and ii) background. To obtain the cancerous area, the extraction of targeted information is very important. To obtain the targeted information, therefore we incorporated the dual CAM at the end to exploit the inter-dependencies among the classes. The architecture of CAM is shown in Fig 2 (bottom). Let us have an original feature map denoted by *X*. Firstly, input feature $X \in R^{H\times W\times C}$ is reshaped into $B \in R^{N\times C}$, whereas $N = H \times W$. Secondly, reshaped *B* and the transpose of *X* are multiplied (matrix multiplication). Thirdly, softmax is applied to the output acquired in the previous step to obtain the attention map $A^{C\times C}$. The channel's inter-dependencies are calculated by using $a_{ji} = \frac{exp(X_i X_j)}{\sum_{i=1}^{C} exp(X_i X_j)}$, such that a_{ji} measures the i_{th} channel's impact on the j_{th} channel. Further, we multiply the $X^{N\times C}$ by $A^{C\times C}$ and reshape the result to $R^{H\times W\times C}$. Lastly, the result is multiplied by a learnable value α and an element-wise summation is performed with the original *X* to obtain the output $Y \in R^{H\times W\times C}$.

The final feature map is calculated by using $Y_j = \alpha \sum_{i=1}^{C}(a_{ji}X_i) + X_j$, which is a weighted sum of features of all channels and an original feature. This results in the modeling of semantic inter-dependencies among channels. Therefore, CAM is applied to the final layers of the proposed CytoNet. Finally, point-wise prediction is obtained by applying softmax to the last feature map.

Experiments

In this section, we describe the whole experiment details starting from network architecture to the evaluation of the proposed method. Moreover, this section is further divided into three sub-sections: implementation details, comparison with state of the art, and architecture ablation. In addition, we explain the performance comparison of the proposed CytoNet on two well-known datasets.

Implementation details

Datasets

To evaluate the proposed method, we used two datasets; i) Our generated thyroid dataset ii) Mendeley Cervical dataset[1]. The comprehensive dataset summary, including the counts of equitably distributed fragments designated for training, validation, and testing, is presented in Table 2. Additionally, Fig. 3 illustrates four randomly chosen samples extracted from datasets pertaining to thyroid and cervical cancers, showcasing diverse backgrounds and colorations. The dataset has been divided into training, validation, and testing parts by the ratio of 80%, 10%, and 10%, respectively. The total number of images in the Mendeley dataset is 963, divided into training, validation, and test data.

Network configuration

In our study, we employed specific hyper-parameters for the training process. The learning rate (lr) was set to 0.001, and a decay learning rate strategy was applied after a step size of 7. The training procedure consisted of 100 epochs, and a batch size of 32 was utilized. The network configuration of CytoNet is shown in Table 4. The proposed network is comprised of two main modules: Efficient-Net, and CAM. First, the input image is given to the Efficient-Net, and then dual CAM is used in parallel to extract target-oriented features. As the Efficient-Net comprises Block 1 to Block 7. Starting from the first Block 1. It comprises mobile convolution layer with a kernel size of 3×3. Similarly, Block2 is composed of two mobile convolution layers with a kernel size of 3×3. Whereas, Block3 consists of two mobile convolution layer with a kernel size of 5×5. Block4 comprises three mobile convolution layer with a size of 3×3, and Block5 consist of three mobile convolution layer with

Types	Dataset	Classifier	ACC	References
PTC	FNAC smears	ANN	85.06	[7]
Breast carcinoma	Histology images	CNN + SVM	88.90	[2]
Lungs carcinoma	Cytological images	CNN	89.30	[5]
Thyroid carcinoma	FNAC images	DT + KNN	90.00	[10]
Thyroid carcinoma	FNAC images	ENN	93.33	[9]
Thyroid carcinoma	FNAC images	SVM	96.70	[8]
PTC	Cytological images	VGG-16	97.66	[6]
Lungs carcinoma	Whole slide images	EfficientNet-B3	98.10	[4]
PTC	Cytological images	Ensemble-Net	99.71	[43]
Proposed	Cytological images	CytoNet	99.00	Own

Table 4. Detailed result analysis of CytoNet with state of the art techniques on our own thyroid dataset. Binary classification results (two classes)

Mendeley dataset				
	Tumor types			
Metrics	Cat2	Cat3	Cat4	Cat5
Precision	1.00	1.00	1.00	0.97
Recall	0.94	1.00	1.00	1.00
F1-score	0.97	1.00	1.00	0.94
SEN	0.941	1.00	1.00	1.00
SPEC	1.00	1.00	1.00	0.988
PPV	1.00	1.00	1.00	0.888
NPV	0.987	1.00	1.00	1.00

Table 5. Detailed result analysis of CytoNet on mendeley dataset (four Classes).

Types	Dataset	Classifier	Classes	ACC
Wang et al. [14]	Histopathology images	ANN	7	88.33
Chandio et al. [44]	Cytopathology images	CNN + SVM	4	74.00
CytoNet (proposed)	Cytopathology images	CytoNet	5	90.00
CytoNet (proposed)	Cytological images	CytoNet	4	99.00

Table 6. Detailed result analysis of CytoNet with state of the art techniques on both datasets: (i) own thyroid dataset (five Classes), and (ii) mendeley dataset (four classes).

Study	Method	Dataset	Classes	Accuracy (%)
Mbaga et al.[45]	SVM	Herlev dataset	7 classes	92.96
Win et al.[46]	SVM, KNN, Boosted trees	SIPaKMeD dataset	2 classes	98.27
		SIPaKMeD dataset	5 classes	94.09
Plissiti et al.[47]	MLP and SVM	SIPaKMeD dataset	5 classes	95.35
Basak et al.[48]	Feature selection and DL	SIPaKMeD dataset	5 classes	97.87
Park et al.[49]	ResNet-50 and SVM	Cervicography images	2 classes	82.00
Tripathi et al.[50]	ResNet-152	SIPaKMeD dataset	5 classes	94.89
Al Mubarak et al.[51]	Fusion based and CNN		4 classes	80.72
Alquran et al.[52]	DL and cascading SVM	Herlev dataset	7 classes	Up to 92
Dhawan et al.[53]	InceptionV3	Kaggle dataset	3 classes	96.10
Huang et al.[54]	ResNet-50V2 and DenseNet-121	Tissue biopsy image dataset	4 classes	95.33
Mulmule and Kanphade[55]	MLP with three kernels and SVM	Benchmark database		97.14
CytoNet (Proposed)	Dual attention	Mendeley	4 classes	99

Table 7. Comparison of CytoNet with previous methods: a summary of studies, methods, datasets, classes, and accuracies.

a size of 5×5. Finally, Block6 and Block7 comprise four mobile convolution layers with the size of 5×5 and 3×3, respectively. The complete block diagram of the CAM is shown in Fig. 2. Whereas, EfficientNeT-B0 and dual CAM are concatenated to each other and the dropout is applied with a linear layer to obtain the final classification results.

Training details

Both of the cancer datasets have class imbalance problems. We have divided both of the datasets into 80% training, 10% validation, and 10% testing. To augment the dataset size, we subdivided the images into smaller patches and trained the model using these patches. Whereas, testing is performed on whole slide image (To evaluate the proposed approach, we used the following evaluation metrics such as accuracy, sensitivity, specificity, precision, recall, F-1 score, positive predictive value (PPV), and negative predictive value (NPV), respectively. We used the cross-entropy loss function for multi-class classification. Whereas, Adam is used as an optimizer with a learning rate of 1e−3. The large training was performed on a NVidia Titan XP GPUs with 64 GB of VRAM.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$SEN = \frac{TP}{TP + FP} \tag{2}$$

$$SPEC = \frac{TP}{TP + FN} \tag{3}$$

$$PPV = \frac{TP}{TP + FP} \tag{4}$$

$$NPV = \frac{TN}{TN + FN} \tag{5}$$

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{8}$$

Comparison with existing algorithms

Evaluation metrics

We evaluated the CytoNet on two datasets (Thyroid, Cervical). We utilized the most commonly used evaluation metrics such as accuracy, sensitivity, specificity, PPV, and NPV for the model evaluation. The accuracy, sensitivity, specificity, PPV, NPV, Precision, Recall and F1 score can be calculated by using Eqs. 1, 2, 3, 4, 5, 6, 7 and 8 respectively. TP, FP, and FN stand for true positive, false positive, and false negative. Whereas PPV stands for positive predictive value and NPV is negative predictive value, respectively.

Experiment results on our own thyroid dataset (binary and multi-class classification)

This experiment was conducted on our own thyroid dataset. Our thyroid dataset consists of 4803 images. We have used 80% of data for training, whereas the rest of the 20% images are used for validation and testing purposes, respectively. We evaluated the proposed model on two well-known datasets. The proposed model produces a comparable result with the existing state-of-the-art approach[43] as shown in Table 4. Whereas, the proposed model outperforms the rest of the existing techniques by a reasonable margin. This is due to the extraction of generic to more detailed features through EfficientNet and CAM as well. The efficient net is capable of obtaining better accuracy compared to conventional CNNs due to compound scaling whereas more localized features are obtained through dual CAM. The accuracy of CytoNet on binary classification is 99.00% as depicted in Table 4. We also evaluate our proposed model on five class datasets as shown in Table 3, whereas the confusion matrix is depicted in 4. Different well-known evaluation metrics like precision, recall, F1-score, sensitivity, specificity, PPV, and NPV are used to judge the multiclass classification performance of the proposed technique.

Experiment results on the mendeley (cervical dataset)

This experiment was conducted on the open-source data known as Mendeley dataset[1]. The repository contains a total of 963 images, categorized into four sets representing the four classes of pre-cancerous and cancerous lesions of cervical cancer according to The Bethesda System standards. These images were captured in 40× magnification using a Leica ICC50 HD microscope and collected from 460 patients using the liquid-based cytology technique. Microscopic examination of cell-level abnormalities facilitates the detection of malignancy or pre-malignant characteristics. We compared the proposed model with well-known methods as shown in Tables 6 and 7. Further, Table 6 depicts the performance of CytoNet on two multi-class datasets (Thyroid, Cervical). The proposed approach outperforms the rest of the techniques by a big margin as shown in Tables 6 and 7. The proposed technique also outperforms the counterpart by achieving 99% accuracy. Whereas detailed evaluation metrics (precision, recall, F1-score, sensitivity, specificity, PPV, NPV) for Medley dataset are depicted in Table 5. The enhanced performance is due to the extraction of shallow to deeper feature extraction. The

Our dataset					
	Tumor types				
Metrics	Cat2	Cat3	Cat4	Cat5	Cat6
Precision	0.88	0.84	0.93	0.82	0.95
Recall	0.89	0.86	0.95	0.86	0.89
F1-score	0.88	0.85	0.94	0.84	0.92
SEN	0.886	0.856	0.952	0.857	0.891
SPEC	0.971	0.974	0.974	0.966	0.985
PPV	0.881	0.842	0.933	0.821	0.951
NPV	0.972	0.977	0.981	0.973	0.966

Table 3. Detailed result analysis of CytoNet on our own thyroid dataset (five classes).

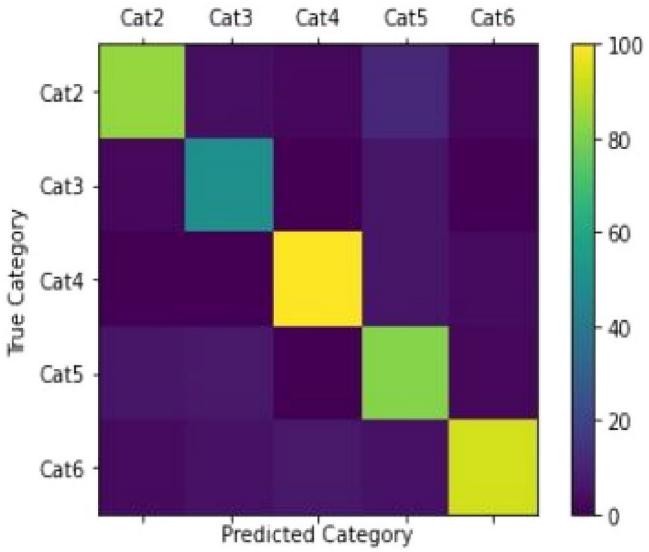

Figure 4. Visualization of confusion matrix of thyroid dataset (five class classification).

combination of compound scaling with the attention mechanism leads to incorporating the generic and more detailed localized features to further increase the accuracy.

Ablation study: evaluating the impact of single and dual channel attention modules (CAM's)

We established the baseline network using EfficientNet-B0 and introduced both single and dual CAM to assess their impact on performance. The experimental comparison between single CAM and dual CAM is presented in Table 8, with a fixed batch size of 32. Our findings demonstrate a noticeable improvement in performance with the integration of dual CAM compared to single CAM. The evaluation metrics in Table 8 provide a detailed insight into the performance comparison between the two networks, highlighting the effectiveness of dual CAM in enhancing overall network performance.

Discussion

The field of cancer diagnosis and screening heavily relies on Computer-Aided Diagnosis (CAD) systems, which have demonstrated limitations in effectively analyzing Whole Slide Images (WSI) and producing accurate results. The primary objective of this research is to develop an automatic classification system capable of accurately distinguishing between various types of carcinomas utilizing deep features. However, the inherent challenges of cluttering, occlusion, and irregular cell distribution in WSI pose significant hurdles in the detection and classification of cancerous cells. Additionally, traditional hand-crafted feature extraction methods have demonstrated reduced accuracy in classification tasks.

In response to these challenges, the proposed CytoNet model has been designed to extract targeted features, aiming to overcome limitations associated with conventional methods. While the model has exhibited improved performance on established datasets, it is essential to acknowledge existing limitations and challenges. These include the necessity for a deeper understanding of the underlying biology and pathology of different cancer subtypes, as well as the requirement for a substantial dataset comprising high-quality images for robust model training and testing.

	Precision	Recall	F1-score	Sensitivity	Specificity	PPV	NPV
EfficientNet-B0 + CAM							
Classes							
Cat2	0.86	0.69	0.77	0.69	0.97	0.85	0.92
Cat3	0.64	0.45	0.53	0.45	0.95	0.63	0.90
Cat4	0.71	0.84	0.77	0.84	0.87	0.70	0.93
Cat5	0.49	0.65	0.56	0.64	0.87	0.49	0.93
Cat6	0.75	0.70	0.73	0.70	0.93	0.75	0.91
Weighted Avg	0.70	0.69	0.69	0.66	0.92	0.68	0.92
EfficientNet-B0 + Dual CAM							
Classes							
Cat2	0.84	0.64	0.73	0.64	0.96	0.83	0.91
Cat3	0.68	0.56	0.61	0.55	0.95	0.68	0.91
Cat4	0.79	0.87	0.83	0.87	0.91	0.79	0.95
Cat5	0.55	0.74	0.63	0.74	0.88	0.55	0.95
Cat6	0.75	0.73	0.74	0.73	0.93	0.75	0.92
Weighted Avg	0.74	0.73	0.73	0.70	0.93	0.72	0.93

Table 8. Ablation study: detailed result analysis of efficient-net with single CAM and efficient-net with dual CAM on our own thyroid dataset (five Classes).

Moving forward, further research is imperative to enhance the accuracy of the model and optimize its network configuration. Additionally, efforts should be directed towards adapting the model to accommodate diverse screening technologies. Addressing these aspects will contribute to the continued advancement of automated cancer diagnosis systems, ultimately facilitating more accurate and efficient cancer screening processes.

Limitations

The CytoNet model, despite its promising capabilities, exhibits certain limitations and challenges that warrant consideration for future research:

- Improvement in the performance of the model: Despite the CytoNet's capabilities, there may be instances where it struggles to accurately predict certain cancer subtypes. This limitation could stem from various factors such as insufficient representation of rare subtypes in the training data, complex tumor characteristics, or inherent variability within cancer types. Addressing this limitation would involve conducting further research to refine the model's algorithms, incorporate additional features or data sources, or implement more advanced machine learning techniques to enhance its predictive accuracy.
- Transfer learning: Given the utilization of a pre-trained EfficientNet model as the backbone of the CytoNet architecture, transfer learning emerges as a pivotal strategy for enhancing the model's capabilities. Transfer learning involves leveraging the knowledge encoded in the pre-trained model, which has been trained on a large and diverse dataset, and fine-tuning it to adapt to the specific characteristics of cytology images and cancer subtypes. By fine-tuning the pre-trained model on a smaller dataset specific to cervical cancer classification, the CytoNet can potentially improve its performance by learning task-specific features and patterns. Therefore, rather than being a limitation, transfer learning represents a valuable opportunity for refining the CytoNet and optimizing its predictive accuracy.
- Network configuration optimization: The architecture and configuration of a neural network play a crucial role in determining its performance. Optimization of the CytoNet's network configuration involves fine-tuning parameters such as the number of layers, the size of each layer, and the choice of activation functions. Adjusting these parameters can have a significant impact on the model's ability to extract meaningful features from the input data and make accurate predictions. By systematically exploring different network configurations and evaluating their performance, researchers can identify optimal settings that maximize the CytoNet's predictive power.
- Integration with different screening technologies: While the CytoNet is specifically designed for cytology images, there is potential to adapt it for use with other screening technologies, such as mammography or CT scans. Each screening modality presents unique challenges and characteristics that must be accounted for in the model's architecture and training process. Adapting the CytoNet to different screening technologies may require modifications to accommodate variations in image resolution, noise levels, and anatomical structures. Additionally, integration with other modalities may necessitate the development of new algorithms or techniques to effectively extract relevant features and interpret the results.Overall, addressing these limitations requires a concerted effort involving multidisciplinary collaboration, advanced algorithm development, and thorough experimental validation. By systematically tackling these challenges, researchers can unlock the full potential of the CytoNet and pave the way for more accurate and reliable cancer diagnosis and screening across a variety of modalities.

Conclusion and future work

In this work, we proposed an end-to-end dual-attention-based automatic prediction of papillary carcinoma using fine needle aspiration cytology. Compared to CNNs, we used the pre-trained efficient net, capable of compound scaling (height, width, depth). Also, dual CAM modules are placed in parallel to obtain the localized targeted features. In this way combination of compound scaling (increases the overall accuracy) and a dual attention module (useful to obtain attention-based features) results in strong feature extraction. Thus CytoNet obtained a comparable performance in terms of accuracy and other metrics such as precision, recall, sensitivity, specificity, and F1-score. In the future, we intend to employ transformers to evaluate the performance of different datasets. In the future, we plan to explore the use of transformers to evaluate the performance of different datasets. Additionally, we aim to investigate the integration of additional modalities, such as clinical data or imaging techniques, to enhance the predictive capabilities of CytoNet. Furthermore, we will explore techniques for model interpretability and visualization to gain insights into the decision-making process of CytoNet. These efforts will contribute to further advancing the accuracy and robustness of our proposed method in clinical practice. In addition to employing transformers for evaluating performance on different datasets, we also intend to delve into the realm of Generative Adversarial Networks (GANs). By leveraging GANs, we aim to explore novel approaches for data augmentation and synthesis, thereby enhancing the robustness and generalization capabilities of our proposed CytoNet model. Through this endeavor, we anticipate further improving the effectiveness and reliability of our automated prediction system for papillary carcinoma using fine needle aspiration cytology.

Data availability

The datasets generated and/or analyzed during the current study are not publicly available due to institutional or regulatory restrictions. Our research institution or governing body has policies in place that limit the public release of certain types of data, such as cancer imaging data, to ensure compliance with ethical standards and data protection regulations. However, the data can be shared by corresponding author with qualified researchers upon request, following the necessary approval processes and adherence to data security and privacy measures.

Received: 11 September 2023; Accepted: 14 October 2024
Published online: 28 October 2024

References

1. Hussain, E. Liquid based cytology pap smear images for multi-class diagnosis of cervical cancer. *Data Brief* (2019).
2. Dabass, M. & Dabass, J. An atrous convolved hybrid seg-net model with residual and attention mechanism for gland detection and segmentation in histopathological images. *Comput. Biol. Med.* **155**, 106690 (2023).
3. Al Nazi, Z., Mashrur, F. R., Islam, M. A. & Saha, S. Fibro-cosanet: Pulmonary fibrosis prognosis prediction using a convolutional self attention network. *Phys. Med. Biol.* **66**, 225013 (2021).
4. Kim, M., Ilyas, N. & Kim, K. Amsaseg: An attention-based multi-scale atrous convolutional neural network for real-time object segmentation from 3d point cloud. *IEEE Access* **9**, 70789–70796 (2021).
5. Ilyas, N. et al. Aaeerp: Advanced auv-aided energy efficient routing protocol for underwater wsns. In *2015 IEEE 29th International Conference on Advanced Information Networking and Applications*, pp. 77–83 (IEEE, 2015).
6. Ilyas, N. *et al.* Extended lifetime based elliptical sink-mobility in depth based routing protocol for uwsns. In *2015 IEEE 29th International Conference on Advanced Information Networking and Applications Workshops*, pp. 297–303 (IEEE, 2015).
7. Khan, A., Han, S., Ilyas, N., Lee, Y.-M. & Lee, B. Cervixformer: A multi-scale swin transformer-based cervical pap-smear wsi classification framework. *Comput. Methods Programs Biomed.* **240**, 107718 (2023).
8. Ilyas, N., Song, Y., Raja, A. & Lee, B. Hybrid-danet: An encoder–decoder based hybrid weights alignment with multi-dilated attention network for automatic brain tumor segmentation. *IEEE Access* **10**, 122658–122669 (2022).
9. Khan, A. & Lee, B. Gene transformer: Transformers for the gene expression-based classification of lung cancer subtypes. arXiv preprint arXiv:2108.11833 (2021).
10. Khan, A., Han, S., Ilyas, N., Lee, Y.-M. & Lee, B. Cervixformer: Transformer-based cervical pap-smear wsi classification framework. Available at SSRN 4266652 .
11. Liu, W. et al. Is the aspect ratio of cells important in deep learning? A robust comparison of deep learning methods for multi-scale cytopathology cell image classification: From convolutional neural networks to visual transformers. *Comput. Biol. Med.* **141**, 105026 (2022).
12. Kim, E., Corte-Real, M. & Baloch, Z. A deep semantic mobile application for thyroid cytopathology. In *Medical imaging 2016: PACS and imaging informatics: next generation and innovations*, vol. 9789, 97890A (International society for optics and photonics, 2016).
13. Guan, Q. et al. Deep convolutional neural network vgg-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images: A pilot study. *J. Cancer* **10**, 4876 (2019).
14. Wang, Y. *et al.* Using deep convolutional neural networks for multi-classification of thyroid tumor by histopathology: A large-scale pilot study. *Ann. Transl. Med.* **7** (2019).
15. Dov, D. *et al.* Thyroid cancer malignancy prediction from whole slide cytopathology images. In *Machine learning for healthcare conference*, pp. 553–570 (PMLR, 2019).
16. Elliott Range, D. D. et al. Application of a machine learning algorithm to predict malignancy in thyroid cytopathology. *Cancer Cytopathol.* **128**, 287–295 (2020).
17. Tao, S. *et al.* Highly efficient follicular segmentation in thyroid cytopathological whole slide image. In *International workshop on health intelligence*, pp. 149–157 (Springer, 2019).
18. Öksüz, C., Urhan, O. & Güllü, M. K. An integrated convolutional neural network with attention guidance for improved performance of medical image classification. *Neural Comput. Appl.* **36**, 2067–2099 (2024).
19. Lubran di Scandalea, M. *et al.* Automatic grading of cervical biopsies by combining full and self-supervision (2022).
20. Tomko, M., Pavliuchenko, M., Pavliuchenko, I., Gordienko, Y. & Stirenko, S. Multi-label classification of cervix types with image size optimization for cervical cancer prescreening by deep learning. In *Inventive computation and information technologies: proceedings of ICICIT 2022*, pp. 885–902 (Springer, 2023).
21. Nugroho, H. A. & Frannita, E. L. Thyroid cancer classification using transfer learning. In *2021 international conference on computer science and engineering (IC2SE)*, vol. 1, pp. 1–5 (IEEE, 2021).
22. Nugroho, H. A., Frannita, E. L. & Hutami, A. H. T. Thyroid nodules categorization based on margin features using deep learning. In *2020 3rd international seminar on research of information technology and intelligent systems (ISRITI)*, pp. 499–504 (IEEE, 2020).

23. Avola, D. et al. Multimodal feature fusion and knowledge-driven learning via experts consult for thyroid nodule classification. *IEEE Trans. Circuits Syst. Video Technol.* **32**, 2527–2534 (2021).
24. Kanavati, F. et al. A deep learning model for cervical cancer screening on liquid-based cytology specimens in whole slide images. *Cancers* **14**, 1159 (2022).
25. Sornapudi, S. et al. Automated cervical digitized histology whole-slide image analysis toolbox. *J. Pathol. Inf.* **12**, 26 (2021).
26. Arifianto, D. & Agoes, A. S. Cervical cancer image classification using cnn transfer learning. In *2nd international seminar of science and applied technology (ISSAT 2021)*, pp. 145–149 (Atlantis Press, 2021).
27. Habtemariam, L. W., Zewde, E. T. & Simegn, G. L. Cervix type and cervical cancer classification system using deep learning techniques. *Med. Dev.* **15**, 163 (2022).
28. Attallah, O. Cervical cancer diagnosis based on multi-domain features using deep learning enhanced by handcrafted descriptors. *Appl. Sci.* **13**, 1916 (2023).
29. Attallah, O. Cercan· net: Cervical cancer classification model via multi-layer feature ensembles of lightweight cnns and transfer learning. *Expert Syst. Appl.* **229**, 120624 (2023).
30. Shinde, S., Kalbhor, M. & Wajire, P. Deepcyto: A hybrid framework for cervical cancer classification by using deep feature fusion of cytology images. *Math. Biosci. Eng.* **19**, 6415–6434 (2022).
31. Liao, X., Huang, Q. & Zheng, X. Necscannet: Novel method for cervical neuroendocrine cancer screening from whole slide images. *Secur. Commun. Netw.* **2021** (2021).
32. Li, T., Feng, M., Wang, Y. & Xu, K. Whole slide images based cervical cancer classification using self-supervised learning and multiple instance learning. In *2021 IEEE 2nd international conference on big data, artificial intelligence and internet of things engineering (ICBAIE)*, pp. 192–195 (IEEE, 2021).
33. Buddhavarapu, V. G. et al. An experimental study on classification of thyroid histopathology images using transfer learning. *Pattern Recogn. Lett.* **140**, 1–9 (2020).
34. Liu, W. et al. Cvm-cervix: A hybrid cervical pap-smear image classification framework using cnn, visual transformer and multilayer perceptron. *Pattern Recogn.* **130**, 108829 (2022).
35. Rahaman, M. M. et al. Deepcervix: A deep learning-based framework for the classification of cervical cells using hybrid deep feature fusion techniques. *Comput. Biol. Med.* **136**, 104649 (2021).
36. Shanthi, P., Hareesha, K. & Kudva, R. Automated detection and classification of cervical cancer using pap smear microscopic images: A comprehensive review and future perspectives. *Eng. Sci.* **19**, 20–41 (2022).
37. Deng, C., Han, D., Feng, M., Lv, Z. & Li, D. Differential diagnostic value of the resnet50, random forest, and ds ensemble models for papillary thyroid carcinoma and other thyroid nodules. *J. Int. Med. Res.* **50**, 03000605221094276 (2022).
38. Sengupta, A., Ye, Y., Wang, R., Liu, C. & Roy, K. Going deeper in spiking neural networks: Vgg and residual architectures. *Front. Neurosci.***13**, 95 (2019).
39. Wu, M., Yan, C., Liu, H., Liu, Q. & Yin, Y. Automatic classification of cervical cancer from cytological images by using convolutional neural network. *Biosci. Rep.* **38** (2018).
40. Akhtar, M., Ali, M. A., Huq, M. & Bakry, M. Fxine-needle aspiration biopsy of papillary thyroid carcinoma: Cytologic, histologic, and ultrastructural correlations. *Diagn. Cytopathol.* **7**, 373–379 (1991).
41. Cibas, E. S. & Ali, S. Z. The 2017 bethesda system for reporting thyroid cytopathology. *Thyroid* **27**, 1341–1346 (2017).
42. Reinhard, E., Adhikhmin, M., Gooch, B. & Shirley, P. Color transfer between images. *IEEE Comput. Gr. Appl.* **21**, 34–41 (2001).
43. Duc, N. T., Lee, Y.-M., Park, J. H. & Lee, B. An ensemble deep learning for automatic prediction of papillary thyroid carcinoma using fine needle aspiration cytology. *Expert Syst. Appl.* **188**, 115927 (2022).
44. Chandio, J. A. & Soomrani, M. A. R. Intelligent diagnostic system for nuclei structure classification of thyroid cancerous and non-cancerous tissues. *Int. J. Adv. Comput. Sci. Appl.* (2017).
45. Mbaga, A. H. & ZhiJun, P. Pap smear images classification for early detection of cervical cancer. *Int. J. Comput. Appl.* **118**, 10–16 (2015).
46. Win, K. P., Kitjaidure, Y., Hamamoto, K. & Myo Aung, T. Computer-assisted screening for cervical cancer using digital image processing of pap smear images. *Appl. Sci.* **10**, 1800 (2020).
47. Plissiti, M. E. *et al.* 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 (IEEE, 2018).
48. Basak, H., Kundu, R., Chakraborty, S. & Das, N. Cervical cytology classification using pca and gwo enhanced deep features selection. *SN Comput. Sci.* **2**, 369 (2021).
49. Park, Y. R. et al. Comparison of machine and deep learning for the classification of cervical cancer based on cervicography images. *Sci. Rep.* **11**, 16143 (2021).
50. Tripathi, A., Arora, A. & Bhan, A. Classification of cervical cancer using deep learning algorithm. In *2021 5th international conference on intelligent computing and control systems (ICICCS)*, pp. 1210–1218 (IEEE, 2021).
51. AlMubarak, H. A. et al. A hybrid deep learning and handcrafted feature approach for cervical cancer digital histology image classification. *Int. J. Healthc. Inf. Syst. Inf.* **14**, 66–87 (2019).
52. Alquran, H. et al. Cervical cancer classification using combined machine learning and deep learning approach. *Comput. Mater. Contin* **72**, 5117–5134 (2022).
53. Dhawan, S., Singh, K. & Arora, M. Cervix image classification for prognosis of cervical cancer using deep neural network with transfer learning. *EAI Endors. Trans. Pervasive Health Technol.* **7** (2021).
54. Huang, P., Tan, X., Chen, C., Lv, X. & Li, Y. Af-senet: Classification of cancer in cervical tissue pathological images based on fusing deep convolution features. *Sensors* **21**, 122 (2020).
55. Mulmule, P. V. & Kanphade, R. D. Supervised classification approach for cervical cancer detection using pap smear images. *Int. J. Medi. Eng. Inf.* **14**, 358–368 (2022).

Acknowledgements

This work was supported by National Research Foundation of Korea funded by the Korean Government (MSIT) under grant RS-2023-00277220.

Author contributions

Naveed Ilyas perform experimentation, Farhat Naseer perform writeup, A.K. help in coding, A.R. give general guidance, Y.M.L., and J.H.P. help to obtain the data, whereas, Boreom Lee mentor the whole project.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to B.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.