EnsembleCAM: Unified Visualization for Explainable Cervical Cancer Identification

Niruthikka Sritharan
Department of Computer Science and Engineering
University of Moratuwa
Moratuwa, Sri Lanka
niruthikka.19@cse.mrt.ac.lk

Nishaanthini Gnanavel
Department of Computer Science and Engineering
University of Moratuwa
Moratuwa, Sri Lanka
nishaanthini.19@cse.mrt.ac.lk

Prathushan Inparaj
Department of Computer Science and Engineering
University of Moratuwa
Moratuwa, Sri Lanka
inparaj.19@cse.mrt.ac.lk

Dulani Meedeniya
Department of Computer Science and Engineering
University of Moratuwa
Moratuwa, Sri Lanka
dulanim@cse.mrt.ac.lk

Pratheepan Yogarajah
School of Computing, Engineering and Intelligent Systems
Ulster University
Londonderry, United Kingdom
p.yogarajah@ulster.ac.uk

***Abstract*—Cervical cancer, which is ranked fourth among cancers affecting women, is highly treatable when detected early through the pap smear test. Deep Learning (DL) models, particularly Convolutional Neural Networks (CNNs), analyze pap smear images, yet their "Black-Box" nature raises transparency concerns in medical diagnostics. This paper introduces a solution named EnsembleCAM to enhance interpretability by unifying visual explanations through the combination of diverse Class Activation Maps (CAMs). Using the Herlev Dataset, we employ data pre-processing, data augmentation techniques, develop an XceptionNet based binary classification model with an accuracy of 89% and apply GradCAM, GradCAM++, Score-CAM, EigenCAM and LayerCAM on this classifier. Then, the novel EnsembleCAM is constructed taking the median of activation maps from the five individual CAM methods. The analysis of activation maps of each CAM method and EnsembleCAM confirmed that in activation maps of EnsembleCAM, higher activation values were more concentrated around the nucleus which is the most important region in indicating cervical malignancy. The evaluation using pixel flipping performance metric also proved that the EnsembleCAM effectively recognises regions vital to the model's decision-making through its steepest drop in the mean prediction score when the pixels in the region contributing most to the model's decision were flipped.**

Keywords*—*class activation maps, ensemble explanations, explainable artificial intelligence, medical image classification, transparent classification visualization

I. Introduction

Cervical cancer ranks as the fourth most prevalent cancer affecting women, yet it stands out as highly treatable when detected early and treated appropriately. The widely adopted and cost-effective screening technique for identifying this form of cancer is the pap smear test [1]. Numerous Deep Learning (DL) models are currently employed for analyzing pap smear images of cervical cells, with Convolutional Neural Networks (CNNs) being prominent for classifying these images into different categories [2]. CNNs have numerous parameters and involve non-linear operations within their convolutional and fully connected layers [3]. Owing to the CNNs' deep architecture, these are regarded as 'Black-Box' models. This opaque nature raises concerns about their workings, despite the remarkable classification performance of these models. In the critical field of medical diagnostics, there is an imperative need for transparent and trustworthy explanations of DL model predictions [4]–[8]. Clear and reliable interpretation of these models not only instils confidence but also aids in identifying decisions made on potentially incorrect information, such as biases or undesirable markings in images.

Class Activation Maps (CAMs) [9] have gained popularity as visual explanation methods for CNNs. Several CAM methods have been proposed to identify the salient regions in an image that influence the underlying prediction model's decisions [10]–[14]. However, as each method uses a different approach, the outputs from these CAM methods have both common and uncommon parts. Owing to the lack of ground truth explanations as well, there is a need to unify the explanations from this multitude of XAI techniques to generate an ensemble explanation that highlights the significant regions in a more precise and reliable manner. Motivated by this, we propose EnsembleCAM, intending to improve the explainability of the classifiers, with the following contributions.

- Propose EnsembleCAM, a novel median-based ensemble method that combines five existing CAMs applied on a cervical cell classification model. To the best of our knowledge, this method has not been proposed in any study in the medical domain.
- Demonstrate that a weighted mean-based ensemble of existing CAMs can be done in a novel manner using the

979-8-3503-7568-8/24/$31.00 ©2024 IEEE

inverse of mean image entropy as the assigned weight.
- Perform qualitative evaluation by analyzing the heatmaps generated by the individual CAM methods and the proposed EnsembleCAM.
- Perform a systematic quantitative evaluation of the component CAM methods and the proposed EnsembleCAM.

II. Background and Related Studies

A. Class Activation Maps

Class Activation Map (CAM) methods serve as visual explanation techniques for CNNs that focus on crucial areas in provided images by leveraging the last layers of convolution containing high-level features of the models. The initial CAM approach necessitates adjustments, including the elimination of the dense layer, incorporation of the Global Average Pooling layer, and subsequent model retraining [9]. Given that the removal of the fully connected layer tends to reduce the model performance, alternative CAM variants have been introduced.

1) GradCAM: Grad-CAM [10] is a method aimed at providing visual interpretations for the decisions produced by CNN-based models. This approach enhances the transparency and interpretability of these models. It operates by utilizing gradients of a chosen target concept that flow into the final convolutional layer, generating a rough localization map that identifies significant areas within the image relevant to predicting the concept. Grad-CAM can be implemented across various CNN model types, such as with fully connected layers, employed for structured outputs and utilized in tasks involving multimodal inputs or reinforcement learning, all without necessitating architectural modifications or retraining.

2) GradCAM++: Grad-CAM++ [11] seeks to enhance the interpretability of CNN predictions, particularly regarding object localization and explaining multiple objects within a single image. It extends the Grad-CAM approach by employing a weighted aggregation of positive partial derivatives from the final convolutional layer feature maps to generate visual explanations for particular class labels. This method has undergone thorough evaluation on established datasets, demonstrating favorable outcomes across tasks such as classification, image captioning, 3D action recognition, and knowledge distillation.

3) Score-CAM: Score-CAM [12] is a post-hoc visual explanation approach designed to unveil the decision-making mechanism of convolutional neural networks. Unlike the previous CAM methods which were reliant on gradients, Score-CAM determines the significance of each activation map by assessing its forward passing score concerning the target class. The final outcome is derived from a linear combination of these weights and activation maps. Score-CAM outperforms previous methods in terms of visual performance and fairness for interpreting the decision-making process. It can be used for both recognition and localization tasks.

4) LayerCAM: LayerCAM [13] efficiently produces class activation maps for all layers by leveraging backward class-specific gradients. Weights for locations with positive gradients are determined by their respective gradients, whereas locations with negative gradients are assigned zero weights. Consequently, the weight is computed for a spatial position (i, j) within the k-th feature map using the ReLU function applied to the respective gradient value. The generation of a class activation map for a specific layer involves the multiplication of each location's activation value in the feature map by its corresponding weight. Finally, the class activation maps for all layers are linearly aggregated along the channel dimension, followed by the application of the ReLU function.

5) Eigen-CAM: Eigen-CAM [14] calculates the principal components of features learned in the convolutional layers of a deep neural network. The image is passed through the neural network model to be explained. Then, the intermediate feature map corresponding to the specified layer of the model is retrieved. Singular Value Decomposition (SVD) is performed on this feature map to obtain singular vectors. Out of these, the first singular vector is selected, multiplied with the feature map, and summed across the channel dimension to form the activation map. Thus, the implementation of Eigen-CAM is independent of class relevance score and can work with any network without the need for modification, training, or backpropagation of parameters across layers.

B. Augmenting or Combining CAMs

Augmented high-order Gradient weighting Class Activation Mapping (augmented grad-CAM++) proposed in [15] combines GradCAM++ activation mappings generated for a set of geometrically augmented images. Following that, bilinear interpolation is used to increase the pixel points in the map in order to enhance the saliency map resolution. Örnek et al. proposed HayCAM [16] a novel visual explanation method that creates an augmented activation map using only the main convolutions in the last convolutional layer with high-class information as selected by Principal Component Analysis.

Normalized outputs from GradCAM, GradCAM++, LayerCAM, and Eigen-CAM are summed and values greater than the fixed threshold of 2 are retained in CodCAM, the ensemble visual explanation proposed by [17]. The ensemble based CAM, MetaCAM, proposed in [18] is also assembled in a way similar to CodCAM. However, instead of using a fixed threshold, they considered adaptive thresholding based on the ROAD (Remove and Debias) performance of MetaCAM.

C. Evaluation of XAI

The key aspects in the evaluation criteria for post hoc explanations in XAI are faithfulness, stability, fairness and interpretability. Faithfulness indicates the accuracy of the XAI method that explains the underlying black-box model. This can be evaluated when the ground truth explanations of the model are available [7], [8], [19]. The stability describes the variation of explanations for perturbing inputs. Drastic changes for small perturbations in the input instance indicate the instability of the XAI method. Fairness is compromised when explanations are more accurate for instances of a given subgroup in the population than for the rest. Interpretability is widely evaluated using human-grounded methods such as forward simulation, that check whether a human can guess the model prediction, given the explanations. Accordingly, several quantitative metrics such as faithfulness correlation, faithfulness estimate, ROAD, Local Lipschitz Estimate, Relative Output Stability and Relative Representation Stability are used to evaluate XAI techniques [20].

III. System Model

A. Process View

This study focuses on developing a method to combine visualizations from various CAM techniques using input images extracted from the Herlev Dataset [21]. After applying data pre-processing and augmentation, we create a binary classification model using XceptionNet [22] to classify cervical cells as either ”normal” or ”abnormal”. Employing XAI techniques such as GradCAM, GradCAM++, Eigen-CAM, Score-CAM, and LayerCAM, we generate class activation maps (CAMs). To enhance interpretability, we calculated the median pixel values across all CAMs, resulting in a new CAM. Performance evaluation is conducted using pixel flipping. The proposed methodology is illustrated in Figure 1.

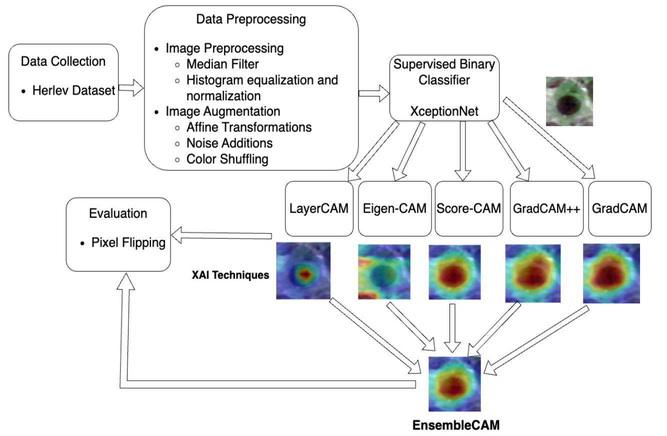


Fig. 1. Overview of the proposed methodology

B. Data Preprocessing

Initially, the images were resized to a consistent 224 x 224 pixels. Then, noise reduction techniques and image enhancement techniques were applied to address the inherent noise and low contrast in Pap smear images [23]. The noise reduction involved using a median filter, while contrast enhancement was achieved through histogram equalization and normalization. These steps increased the contrast, aiding in the extraction of important information from the images. Figure 2 shows a sample of the implemented data preprocessing techniques.

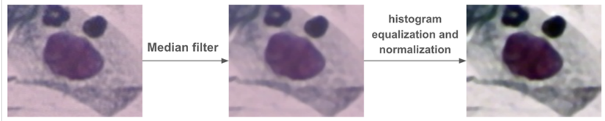
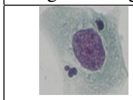
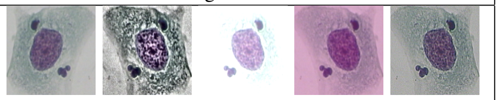


Fig. 2. Data Preprocessing

C. Data Augmentation

Data augmentation is vital to expand and diversify the training dataset. In this study, an elaborate data augmentation pipeline was implemented to generate about 8 augmentations for each cervical cell image using randomly chosen augmentation techniques from a defined set of augmentations. Some of the applied augmentation techniques were translation, scaling, rotations, distortions, gamma adjustments for brightness, Gaussian noise addition, blurring, image sharpening and colour-related transformations. Table I shows a sample of the augmentations.

TABLE I. Sample of Data Augmentations

Original Image	Augmentations
	

D. Classification Model

For cervical cell classification, XceptionNet was utilized as a feature extractor with pre-trained weights on the ImageNet dataset. XceptionNet is a deep convolutional neural network which enhances parameter utilization for image classification by incorporating depthwise separable convolutions, redefining traditional Inception modules [22]. The architecture employs depthwise convolutions for spatial information and pointwise convolutions for cross-channel correlations. With 36 convolutional layers, including 14 residual layers with bypass connections, XceptionNet is organized into three flows: Entry Flow, Middle Flow, and Exit Flow.

The model extended the base architecture by appending a Global Average Pooling layer to reduce spatial dimensions, followed by Batch Normalization for improved training stability. Subsequently, a densely connected layer with 256 units and ReLU activation was introduced, incorporating regularization to prevent overfitting. A dropout layer with a rate of 0.45 was then employed for further regularization. The final layer consisted of a dense softmax layer with the number of output classes. The base XceptionNet layers were frozen to retain the pre-trained knowledge during fine-tuning and the model was compiled using the Adam optimizer and categorical cross-entropy as the loss function. Figure 3 shows the process of the cervical cell classification model.

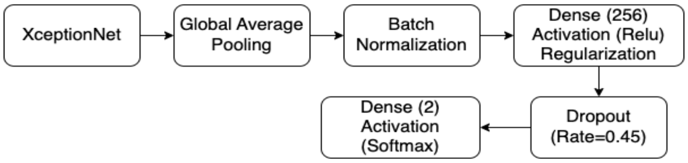

XceptionNet
Global Average Pooling
Batch Normalization
Dense (256) Activation (Relu) Regularization
Dropout (Rate=0.45)
Dense (2) Activation (Softmax)

Fig. 3. Process of the cervical cell classification model

E. Ensemble Explanation Generation

The CAM methods namely, GradCAM, GradCAM++, Score-CAM, Eigen-CAM and LayerCAM were applied on the activation layer preceding the Global Average Pooling layer of the above-mentioned XceptionNet classification model. The activation map returned from each CAM method was normalized to have values in the range [0, 1] and resized to the original height and width of the image. Each CAM method adopts a distinct strategy; for instance, GradCAM relies on first-order gradients, GradCAM++ utilizes second-order gradients, Score-CAM uses a weighted linear combination of activations, Eigen-CAM relies on the initial components of the activations and LayerCAM involves different layers. Consequently, the activation maps generated by these methods exhibited variations. Given the absence of ground-truth explanations for cervical cell images, an ensemble activation map proved useful in highlighting the significant regions of the image influencing the model's decision.

1) Mean Based Ensemble: The first approach to combine these activation maps from the 5 CAM methods was to take the mean of the individual activation maps. Equation (1) demonstrates the aggregation of individual activation maps by computing their mean. Here, $[G_{ij}]_{m\times n}$, $[GP_{ij}]_{m\times n}$, $[S_{ij}]_{m\times n}$, $[E_{ij}]_{m\times n}$ and $[L_{ij}]_{m\times n}$ indicate the activation maps obtained from GradCAM, GradCAM++, Score-CAM, Eigen-CAM and LayerCAM respectively. $[M1_{ij}]_{m\times n}$ is used to denote the mean-based ensemble activation map.

$$[M1_{ij}]_{m\times n} = \frac{[G_{ij}]_{m\times n}+[GP_{ij}]_{m\times n}+[S_{ij}]_{m\times n}+[E_{ij}]_{m\times n}+[L_{ij}]_{m\times n}}{5} \quad (1)$$

Though this is a simple and intuitive method, if one of the activation maps has significantly lower or higher values compared to the others, mean's sensitivity to outliers can disproportionately influence the final ensemble map. This will lead to a skewed representation of the combined saliency information. Though noise can be reduced via averaging, boundaries between salient and non-salient regions may also be blurred reducing the precision of this ensemble explanation. To address these limitations, the weighted mean was next considered to combine the activation maps.

2) Weighted Mean Based Ensemble: A quantitative measure of the CAMs is required to determine the weight that can be assigned to each activation map from the 5 CAM methods. Image entropy is used for this task, which measures the disorder or randomness within the pixel values of an image. A higher entropy indicates that the explanation is overly complex, capturing both relevant and irrelevant image features. However, a lower entropy value suggests that the explanation is more focused and less random, potentially highlighting only the significant and relevant regions of the image.

This metric was selected rather than any other quantitative metric since it can assist in gauging all 4 key aspects of XAI evaluation. A lower entropy value only emphasizes the key features in the image that align with the model's decision, thus representing a higher faithfulness. Entropy, by capturing the randomness, assesses stability by suggesting that lower entropy corresponds to more stable explanations. Additionally, if entropy is high, it hints that the XAI method is capturing irrelevant features that could potentially introduce bias. Also, a more focused and less complex explanation yielding a lower entropy value is bound to provide higher interpretability for humans. Therefore, the CAM method that produces explanations having the least entropy should be given the highest weight when creating the ensemble activation map.

Accordingly, image entropy was computed for the explanation generated by each CAM method for each image in the test set. Then, the entropy values were averaged across all the images and a mean image entropy value was obtained for each CAM method. The inverse of the corresponding mean image entropy was utilized as the weight of each CAM method to produce the weighted mean-based ensemble activation map. Construction of Weighted Mean-Based Ensemble explained step-by-step through the progression of (2), (3), and (4). Here, $\frac{1}{e_G}$, $\frac{1}{e_{GP}}$, $\frac{1}{e_S}$, $\frac{1}{e_E}$ and $\frac{1}{e_L}$ indicate the inverses of the mean image entropy values computed for GradCAM, GradCAM++, Score-CAM, Eigen-CAM and LayerCAM respectively. $[M2_{ij}]_{m\times n}$ denotes the weighted mean based ensemble activation map.

However, the high subjectivity to the weight assignment makes this method less robust. Variations in dataset characteristics or model outputs cause significant changes to the weights assigned to each CAM method, leading to the absence of a unified criterion for weight determination.

$$\begin{aligned}[T_{ij}]_{m\times n} = {} & \frac{1}{e_G}[G_{ij}]_{m\times n} + \frac{1}{e_{GP}}[GP_{ij}]_{m\times n} \\ & + \frac{1}{e_S}[S_{ij}]_{m\times n} + \frac{1}{e_E}[E_{ij}]_{m\times n} + \frac{1}{e_L}[L_{ij}]_{m\times n}\end{aligned} \quad (2)$$

$$w_{total} = \frac{1}{e_G} + \frac{1}{e_{GP}} + \frac{1}{e_S} + \frac{1}{e_E} + \frac{1}{e_L} \quad (3)$$

$$[M2_{ij}]_{m\times n} = \frac{[T_{ij}]_{m\times n}}{w_{total}} \quad (4)$$

3) EnsembleCAM: Median Based Ensemble: In this novel approach the activation maps from the 5 CAM methods were assembled using median, which is the middle value in the sorted list of activation values. This intuitive method offers many advantages compared to the previous approaches. It is less sensitive to extreme values in any component activation maps, thus preventing skewed interpretations. It also tends to be less prone to fluctuations, ensuring more robustness. The EnsembleCAM algorithm is demonstrated in Algorithm 1.

Algorithm 1 EnsembleCAM (proposed)

```
Require: GradCAM, GradCAM++, Score-CAM, LayerCAM, Eigen-CAM
$gCAM \leftarrow$ GetGradCAM
$gpCAM \leftarrow$ GetGradCAMPlusPlus
$sCAM \leftarrow$ GetScoreCAM
$lCAM \leftarrow$ GetLayerCAM
$eCAM \leftarrow$ getEigenCAM
$cams \leftarrow [gCAM, gpCAM, sCAM, lCAM, eCAM]$
$processedCAMS \leftarrow []$
for each cam in cams do
    $x \leftarrow \text{normalize}(cam)$
    $x \leftarrow \text{resize}(x, \text{width}, \text{height})$
    $processedCams \leftarrow processedCams + [x]$
end for
$EnsembleCAM \leftarrow$ median(processedCams)
```

F. Evaluation Metrics

Pixel flipping performance metric is used for the evaluation of this study. This metric is grounded in the pixel-flipping methodology [24], which involves a systematic alteration of individual pixels within an image, commencing from the most relevant ones as per the provided explanation and advancing towards the less relevant ones. The main aim of this approach is to simulate the influence of perturbations on the image by measuring the decline in prediction scores as pixels undergo progressive flipping. Moreover, it observes the consequential impact on the explanation produced by the XAI method. Analyzing the effect of each flipped pixel on the explanation contributes to the assessment of the explanation's reliability and sensitivity to pixel-level changes. A higher degree of stability, where minor pixel adjustments yield consistent explanations, signifies a more reliable and robust XAI methodology.

IV. Results and Discussion

A. Performance of Classification Model

We used a batch size of 16, a learning rate of 0.001 and trained the cervical cell classification model for 20 epochs. Our evaluation metrics included accuracy, precision, recall, and F1 score. The resulting model achieved an accuracy of 89%, with a precision score of 0.9, a recall score of 0.89 and an F1 score of 0.88. From these performance metrics, it is evident that the model has performed well in cervical cell classification.

B. Performance of EnsembleCAM

As a qualitative evaluation method, a sample of cervical cell image from each of the normal and abnormal class are selected to visually compare the performance of the other five CAM methods with the proposed EnsembleCAM. Heatmaps were generated for each of the 5 CAM methods and EnsembleCAM and overlaid on the original images. The original and the overlaid images are shown in Table II. In the EnsembleCAM activation maps, the higher activation values, which are shown by the red-coloured region, are more concentrated around the nucleus, which is the primary indicator of malignancy in cervical cells. Further, the highlighted regions in EnsembleCAM are more distinct and well-defined compared to the corresponding regions in the other CAM methods.

As a quantitative evaluation, we evaluated each of the 5 CAM methods and EnsembleCAM, utilizing the pixel-flipping performance metric. A graph was generated to illustrate the mean prediction confidence score for cervical cell images as pixels in the region contributing most to the model's decision were flipped at varying percentages. The resulting graph is shown in Figure 4. Analysis of Figure 4 reveals that the EnsembleCAM curve has the steepest drop compared to the other 5 CAM methods. This signifies that EnsembleCAM effectively identifies regions crucial to the model's decision, as substantial perturbations in that region result in significant shifts in the model's decision and prediction confidence scores.

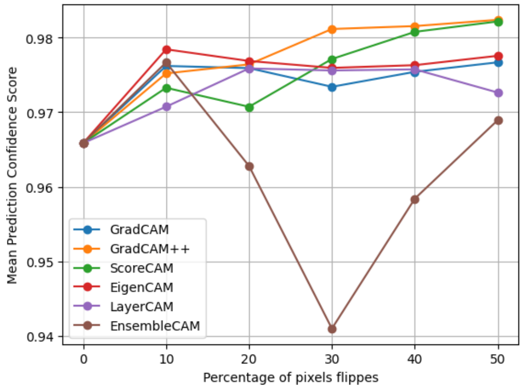
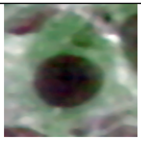
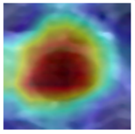
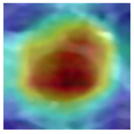
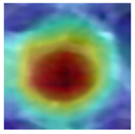
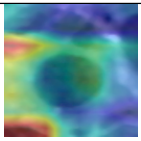
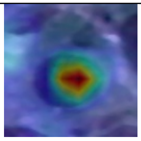
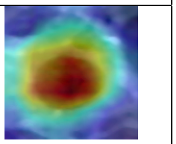
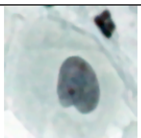
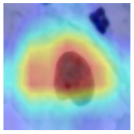
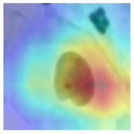
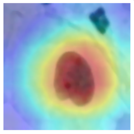
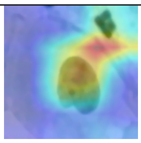
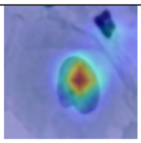
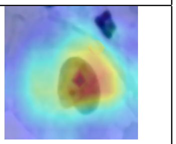


Fig. 4. Pixel Flipping Graph

V. Conclusion

This study presented an approach to visually explain the regions-of-interests for cervical cancer image classification. We trained an XceptionNet model that classifies cervical cell images into normal and abnormal. However, without visual explanations, model decisions are not fully reliable since medical specialists would like to see why the XceptionNet model decided a cervical cell image as malignant and the salient regions in the image influencing this decision. Therefore, we applied 5 different CAM methods on the activation layer preceding the global max pooling layer of the XceptionNet model to generate

TABLE II. Comparison of individual CAMs and EnsembleCAM

	Original	GradCAM	GradCAM++	Score-CAM	Eigen-CAM	LayerCAM	**EnsembleCAM**
Normal							
Abnormal							

activation maps as explanations. Owing to the differences in the techniques adopted by each CAM method, the generated activation maps have both common and uncommon areas. Given the absence of ground-truth explanations as well, we proposed a new ensemble visual explanation method named EnsembleCAM. The examination of activation maps from 5 component CAM methods and EnsembleCAM, affirms that EnsembleCAM exhibits higher activation values concentrated around the crucial nucleus region, a key indicator of cervical malignancy. Pixel flipping analysis further demonstrates that EnsembleCAM effectively identifies pivotal decision-making regions by showing the most significant decline in mean prediction scores when key pixels are flipped. Thus, the qualitative and quantitative evaluations show that the ensemble explanations generated by EnsembleCAM explain the decisions of the underlying blackbox classification model better than the component CAM methods.

References

[1] M. M. Rahaman, C. Li, Y. Yao, F. Kulwa, X. Wu, X. Li, and Q. Wang, "Deepcervix: A deep learning-based framework for the classification of cervical cells using hybrid deep feature fusion techniques," *Computers in Biology and Medicine*, vol. 136, p. 104649, 2021.

[2] C. Yang, L.-h. Qin, Y.-e. Xie, and J.-y. Liao, "Deep learning in CT image segmentation of cervical cancer: a systematic review and meta-analysis," *Radiation Oncology*, vol. 17, no. 1, p. 175, 2022.

[3] D. Meedeniya, *Deep Learning: A Beginners' Guide*. CRC Press LLC, 2023. [Online]. Available: www.routledge.com/9781032473246

[4] S. Wickramanayake, S. Rasnayaka, M. Gamage, D. Meedeniya, and I. Perera, "Chapter one - explainable artificial intelligence for enhanced living environments: A study on user perspective," in *Internet of Things: Architectures for Enhanced Living Environments*, ser. Advances in Computers, G. Marques, Ed. Elsevier, 2024, vol. 133, pp. 1–32.

[5] S. Dasanayaka, V. Shantha, S. Silva, T. Ambegoda, and D. Meedeniya, "Interpretable machine learning for brain tumor analysis using MRI," in *Proceedings of the 2nd International Conference on Advanced Research in Computing (ICARC)*, Belihuloya, Sri Lanka, 2022, pp. 212–217.

[6] N. Gnanavel, P. Inparaj, N. Sritharan, D. Meedeniya, and P. Yogarajah, "Interpretable cervical cell classification: A comparative analysis," in *Proceedings of the 4th International Conference on Advanced Research in Computing (ICARC)*. Belihuloya, Sri Lanka: IEEE, 2024, pp. 1–6.

[7] T. Shyamalee, D. Meedeniya, G. Lim, and M. Karunarathne, "Automated tool support for glaucoma identification with explainability using fundus images," *IEEE Access*, vol. 12, pp. 17 290–17 307, 2024.

[8] L. Gamage, U. Isuranga, D. Meedeniya, S. De Silva, and P. Yogarajah, "Melanoma skin cancer identification with explainability utilizing mask guided technique," *Electronics*, vol. 13, no. 4, 2024.

[9] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, Las Vegas, USA, 2016, pp. 2921–2929.

[10] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, Venice, Italy, 2017, pp. 618–626.

[11] A. Chattopadhay, A. Sarkar, P. Howlader, and V. N. Balasubramanian, "Grad-CAM++: Generalized gradient-based visual explanations for deep convolutional networks," in *IEEE winter conference on applications of computer vision (WACV)*. Lake Tahoe, USA: IEEE, 2018, pp. 839–847.

[12] H. Wang, Z. Wang, M. Du, F. Yang, Z. Zhang, S. Ding, P. Mardziel, and X. Hu, "Score-CAM: Score-weighted visual explanations for convolutional neural networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, Seattle, USA., 2020, pp. 24–25.

[13] P.-T. Jiang, C.-B. Zhang, Q. Hou, M.-M. Cheng, and Y. Wei, "LayerCAM: Exploring hierarchical class activation maps for localization," *IEEE Transactions on Image Processing*, vol. 30, pp. 5875–5888, 2021.

[14] M. B. Muhammad and M. Yeasin, "Eigen-CAM: Class activation map using principal components," in *International joint conference on neural networks (IJCNN)*. Glasgow, United Kingdom: IEEE, 2020, pp. 1–7.

[15] Y. Gao, J. Liu, W. Li, M. Hou, Y. Li, and H. Zhao, "Augmented Grad-CAM++: Super-resolution saliency maps for visual interpretation of deep neural network," *Electronics*, vol. 12, no. 23, p. 4846, 2023.

[16] A. H. Örnek and M. Ceylan, "HayCAM: A novel visual explanation for deep convolutional neural networks," *Traitement du Signal*, vol. 39, no. 5, pp. 1711–1719, 2022.

[17] A. H. Ornek and M. Ceylan, "CodCAM: A new ensemble visual explanation for classification of medical thermal images," *Quantitative InfraRed Thermography Journal*, pp. 1–25, 2023.

[18] E. Kaczmarek, O. X. Miguel, A. C. Bowie, R. Ducharme, A. L. Dingwall-Harvey, S. Hawken, C. M. Armour, M. C. Walker, and K. Dick, "MetaCAM: Ensemble-based class activation map," *arXiv preprint arXiv:2307.16863*, 2023.

[19] S. Dasanayaka, V. Shantha, S. Silva, D. Meedeniya, and T. Ambegoda, "Interpretable machine learning for brain tumour analysis using mri and whole slide images," *Software Impacts*, vol. 13, p. 100340, 2022.

[20] A. Hedström, L. Weber, D. Krakowczyk, D. Bareeva, F. Motzkus, W. Samek, S. Lapuschkin, and M. M. M. Höhne, "Quantus: An explainable ai toolkit for responsible evaluation of neural network explanations and beyond," *Journal of Machine Learning Research*, vol. 24, no. 34, pp. 1–11, 2023.

[21] J. Jantzen, J. Norup, G. Dounias, and B. Bjerregaard, "Pap-smear benchmark data for pattern classification," *Nature inspired smart information systems (NiSIS)*, pp. 1–9, 2005.

[22] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251–1258.

[23] H. Alquran, M. Alsalatie, W. A. Mustafa, R. A. Abdi, and A. R. Ismail, "Cervical net: A novel cervical cancer classification using feature fusion," *Bioengineering*, vol. 9, no. 10, p. 578, 2022.

[24] J. Kauffmann, K.-R. Müller, and G. Montavon, "Towards explaining anomalies: a deep taylor decomposition of one-class models," *Pattern Recognition*, vol. 101, p. 107198, 2020.