Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 7, No. 3, July 2025, pp: 567-580 e-ISSN: 2656-8632

RESEARCH ARTICLE OPEN ACCESS

Grad-CAM Based Visualization for Interpretable Lung Cancer Categorization Using Deep CNN Models

Rashmi Mothkur[1], Pullagura Soubhagyalakshmi[2], and Swetha C. B.[1]

[1] Alliance School of Advanced Computing, Alliance University, Bangalore, Karnataka, India
[2] Department of Computer Science and Engineering, K.S.Institute of Technology, Bangalore, Karnataka, India

Corresponding author: Rashmi Mothkur (e-mail: rashmimothkur@gmail.com)

ABSTRACT The Grad-CAM (Gradient-weighted Class Activation Mapping) technique has loomed as a crucial tool for elucidating deep learning models, particularly convolutional neural networks (CNNs), by visually accentuating the regions of input images that accord most to a model's predictions. In the context of lung cancer histopathological image classification, this approach provides discernment into the decision-making process of models like InceptionV3, XceptionNet, and VGG19. These CNN architectures, renowned for their high performance in image categorization tasks, can be leveraged for automated diagnosis of lung cancer from histopathological images. By applying Grad-CAM to these models, heatmaps can be generated that divulge the areas of the tissue samples most influential in categorizing the images as lung adenocarcinomas, squamous cell carcinoma, and benign patches. This technique allows for the visualization of the network's focus on specific regions, such as cancerous cells or abnormal tissue structures, which may otherwise be difficult to explicate. Using pre-trained models fine-tuned for the task, the Grad-CAM method assesses the gradients of the target class concerning the final convolutional layer, generating a heatmap that can be overlaid on the input image. The results of Grad-CAM for InceptionV3, XceptionNet, and VGG19 offer distinct insights, as each model has unique characteristics. InceptionV3 pivots on multi-scale features, XceptionNet apprehend deeper patterns with separable convolutions, and VGG19 emphasizes simpler, more global attributes. By justaposing the heatmaps generated by each architecture, one can assess the model's focus areas, facilitating better comprehension and certainty in the model's prophecy, crucial for clinical applications. Ultimately, the Grad-CAM approach not only intensify model transparency but also aids in ameliorating the interpretability of lung cancer diagnosis in histopathological image categorization.

Keyword Explainable AI, Grad-CAM, InceptionV3, VGG19, XceptionNet.

I. Introduction

Rapid advances in artificial intelligence have transformed the healthcare business, allowing for the creation of complex algorithms capable of processing massive volumes of data, recognizing patterns, and generating predictions with unparalleled precision. However, the increasing use of these AI-powered systems has prompted questions about their interpretability and transparency. A major obstacle in the clinical implementation of AI algorithms is the lack of trust and understanding among healthcare providers and patients, who frequently view these systems as incomprehensible "black boxes" whose internal structures and decision-making processes are not obvious or easily understandable. Addressing this issue, the concept of Explainable Artificial Intelligence has emerged as a prospective solution. Explainable AI aims to make AI algorithms more transparent and comprehensible, allowing clinicians and patients to better comprehend the rationale behind the predictions and decisions made by these systems[1]. This is particularly critical in the healthcare domain, where decisions can have profound implications for patient outcomes and safety. The significance of interpretability and limpidness in AI-powered healthcare systems has been widely recognized. Clinicians and patients are more likely to trust and adopt AI algorithms that can provide clear elucidation for their outputs, thereby strengthening the overall acceptance and integration of these technologies into clinical practice [2].

A. Background

The rapid furtherance of artificial intelligence (AI) in healthcare has transfigured diagnostic processes,

Manuscript received January 16, 2025; Accepted April 20, 2025; date of publication May 5, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.690

treatment planning, and patient management. However, the increasing complexity of AI algorithms poses noteworthy provocation in terms of transparency and interpretability. As healthcare professionals rely on these technologies to make critical decisions, the need for explainable AI (XAI) has become paramount[3]. Explainable AI (XAI) refers to a set of techniques that make AI decisions interpretable and understandable to humans. In the healthcare domain, where AI-driven decisions can significantly impact patient outcomes, interpretability is essential to ensure trust among healthcare professionals and patients. XAI aims to provide insights into how AI models arrive at their prognosis, thereby enhancing trust and facilitating better collaboration between clinicians and AI systems.

Given the interpretability gaps in current XAI approaches, there is a growing need for enhanced frameworks that balance computational efficiency, stability, and clinical relevance. Future research should focus on hybrid models that integrate multiple interpretability techniques, ensuring robust and clinically meaningful explanations. Additionally, user-centric evaluation, incorporating feedback from radiologists, is essential to refining AI-driven diagnostic tools for lung cancer.

B. Rationale and Objectives

Improving interpretability in AI-driven medical imaging can significantly impact clinical workflows and patient outcomes. When AI models provide clear, reliable explanations for their decisions, radiologists and oncologists can integrate AI-assisted diagnostics into their workflow with greater confidence. This can lead to:

1. Faster and More Accurate Diagnoses: Enhanced interpretability allows AI models to highlight key diagnostic features, reducing ambiguity and expediting decision-making.
2. Improved Treatment Planning: Clinicians can better understand AI-driven insights, leading to personalized and precise treatment strategies for lung cancer patients.
3. Increased Trust and Adoption: Transparent AI models encourage greater trust among healthcare professionals, facilitating wider adoption and integration into daily practice.
4. Reduced Diagnostic Errors: By making AI decisions interpretable, clinicians can identify potential model biases or errors, ensuring safer and more reliable patient care.

The ramifications of opaque decision-making can be severe in the high-stakes healthcare industry. For example, knowing the reasoning behind AI-generated suggestions is crucial for guaranteeing patient safety and adherence to clinical guidelines in fields like radiology, cancer, and personalized medicine[4]. Additionally, regulatory agencies are stressing the need for models that medical practitioners can examine and comprehend, highlighting the need of interpretability in AI applications. The key contributions of the proposed work are as follows:

1. This paper aims to compare the effectiveness of various CNN architectures in lung cancer image classification and to evaluate their interpretability through Grad-CAM.
2. It employs InceptionV3, VGG-19 and XceptionNet into categorization of lung adenocarcinomas, lung squamous cell carcinoma, and benign patches. Various rudimentary model evaluation metrics are computed. These metrics aids in selection of the unrivalled model amongst the various models trained.
3. Utilize Grad-CAM to recognize and visualize the regions of histopathological images that contribute most significantly to the prognosis made by CNN architectures.

II. Related Works

The latest XAI methods used in medical imaging are covered by Singh, Sengupta et al.[5] , with an emphasis on categorizing lung cancer from CT and X-ray images. The authors examine numerous approaches for rendering deep learning models interpretable, including Grad-CAM and LIME, and offer guidance on how these techniques can assist medical professionals in interpreting AI-driven judgments, particularly in diagnosing lung cancer.

The use of XAI methods such as SHAP and LIME with convolutional neural networks (CNNs) for the categorization of lung cancer from CT and X-ray images is examined by Ganaie, Muhammad et al[6]. The study highlights how XAI techniques may detect important picture elements that affect the prediction result, including tumors or nodules. The authors suggest a framework that takes interpretability and classification accuracy into account.

According to Kai Gao, Hui Shen et al [7], the primary approach to resolving the interpretability issue with deep networks is the display of deep models. One popular technique for visualizing deep models is the Class Activation Map (CAM). However, because CAMs only visualize the final layer, their resolution is limited. In this work, we combined Dense Net and CAM to create a new convolutional network called Dense-CAM. We then visualized the entire network to produce a deep model visualization that is more reliable and accurate. Using almost 6000 samples, the network was evaluated on the gender categorization issue and achieved a 92.93% accuracy rate. Brain regions with significant dissimilarity between men and women are found with the proposed method, which can be used for future brain imaging studies.

Manuscript received January 16, 2025; Accepted April 20, 2025; date of publication May 5, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.690

Teramoto et al [8]used a deep convolutional neural network (DCNN), a popular deep learning method, to create an automated categorization scheme for lung tumors shown in microscopic images. Three convolutional layers, three pooling layers, and two penultimate linked layers constitute the classification phase of DCNN. The DCNN was trained utilizing a graphics processing unit and our original database in assessment studies. Prior to being augmented by rotation, flipping, and filtering to minimize overfitting, microscopic images were initially resized and resampled to generate images with a resolution of 256 × 256 pixels. The employed method was used to evaluate the probability of three forms of cancer, and threefold cross validation was used to assess the categorization accuracy. About 71% of the images in the results were correctly categorized, which is analogous to the accuracy of pathologists and cytotechnologists. The developed method is efficient in categorizing lung tumors from microscopic images.
Fig. 1 depicts various XAI methods. The following section discusses the general categories of XAI techniques and their contribution in making AI systems more comprehensible and transparent.

A. Transparent Models

These models are easy to use and can potentially place into practice quickly. Simple computations that even humans can perform make up the algorithms. As a result, these models can be explained, and people may readily comprehend how they reach a particular conclusion[9]. Examples of transparent models include linear regression, where resultants are determined by weighted sums of input attributes, and logistic regression, which assigns probabilities to various outcomes based on input data. Decision trees are simple rule-based structures that relegate decisions into a series of conditions, making them easy to visualize. Similarly, rule-based models explicitly define IF-THEN rules, often used in expert systems and business logic. The k-nearest neighbors (k-NN) algorithm, though less decipherable in high-dimensional spaces, is intuitive in lower dimensions, as prognosis are based on similarity to known data points. Naïve Bayes classifiers also offer transparency by relying on probabilistic distributions and conditional independence assumptions.

B. Model Agnostic

After a model has been learned to explicate its forecasts, post-hoc procedures are used in Explainable AI (XAI). In Explainable AI (XAI), "model agnosticism" refers to strategies and tactics that may be used with any machine learning model, irrespective of its intricacy or underlying architecture[10]. It can be used with different models without requiring significant changes or retraining. Examines a particular prediction by developing a local, interpretable model around it, demonstrating the impact of altering input characteristics on the result. A type of machine learning where the model makes prognosis based on specific instances of the training data rather than learning a general model. In the context of Explainable AI (XAI), IBL methods can provide interpretable predictions because they often rely on the similarity between instances. Table 1 depicts the various works on XAI techniques

Table 1. CAM and LIME Techniques

Authors	Technique	Advantages
Xiao et al[11]	Improved Grad CAM	Enhances the visualization of critical image regions.
Selvaraju et al[12]	Grad CAM	Highlights important image areas for concept prediction
Chen et al [13]	C CAM Method	Improves foreground and background boundaries in segmentation
Vinogradova et al [14]	SEG-GRAD-CAM	Provides local representations for segmentations using gradients
Ribeiro et al [15]	LIME	Provides interpretability for any classifier prediction
Schiavon et al[16]	GRAD-CAM	Predicts discrete subtypes of brain tumor
Ahsan et al [17]	LIME	Provides a better understanding of the features in CT /Xray Images on COVID 19

C. Model Specific

Model-specific XAI techniques are tailored to provide explanations for certain types of machine learning models. These methods anchorage the unique characteristics and structures of the models to deliver more insightful and often more efficient explanations. These methods can be computationally more efficient than model-agnostic approaches, especially for simpler models[18].It makes the decision-making process easy to grasp by presenting facts in an understandable, human-readable manner. A set of conditional statements (if-then rules) that can be linked to data aspects serve as the basis for decision-making. It highlights regions of an input that accords most to the model's output, allowing users to see which features (or parts of the image) influenced the prediction the most. Saliency maps are typically generated using the gradients of the output with regard to the input[19]. By calculating these gradients, the maps show how

Manuscript received January 16, 2025; Accepted April 20, 2025; date of publication May 5, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.690

sensitive the model's prediction is to changes in each pixel.

D. Inceptionnet (Googlenet)

Google unveiled InceptionNet, a deep convolutional neural network that uses the Inception module, in 2014. By using convolutions of varying sizes (1x1, 3x3, 5x5) in parallel, this module enables a more effective use of computational resources while capturing a range of feature scales. It greatly reduces overfitting by using global average pooling in place of fully connected layers[20]. With more than 20 layers, the architecture is deep and includes dimensionality reduction strategies like 1x1 convolutions to aid in calculation speedup[21]. Particularly in tasks requiring multi-scale feature extraction, such as tumor identification, lesion segmentation, and organ categorization in medical imaging like CT scans, MRIs, and X-rays, InceptionNet has demonstrated great promise in the medical field. An algorithm developed by Pranav Rajpurkar et al[22] identify pneumonia from chest X-rays more accurately than trained radiologists. With over 100,000 frontal-view X-ray images of 14 diseases, ChestX-ray14 is presently the widely used open source chest X-ray dataset.The work justapoxed CheXNet's performance to radiologist’s prognosis using a test set annotated by four active academic radiologists. On the F1 metric, the work evaluated that CheXNet accomplishes better than the average radiologist.

E. Xceptionnet

François Chollet[23] presented the Xception architecture in 2017, which enhances InceptionNet by substituting depthwise separable convolutions for the usual convolutions. This method divides the convolution process into two stages: pointwise convolutions (1x1 convolutions across the output channels) and depthwise convolutions.This preserves great representational power while lowering the number of parameters.

B. Uma Maheswari et al [24]proposed model consisting of four convolution-MaxPooling layers with different hyperparameters that were ameliorated for feasible performance using a Bayesian optimization technique. The model was delineated with a apex categorization evaluation metrices of accuracy with 0.95. In addition, the receiver operating characteristic (ROC) curve for the proposed shallow-CNN showed a apex area under the curve value of 0.976. Moreover, they have incorporated class activation maps (CAM) and Local Interpretable Model-agnostic Explanations (LIME), explainer systems for estimating the limpidness and explainability of the model in contrast to a existing pre-trained neural net such as the DenseNet.

F. VGG-19

The Visual Geometry Group (VGG) introduced the VGG-19, a deep convolutional neural network renowned for its straightforward yet profound architecture, in 2014. Three fully connected layers and sixteen convolutional layers make up its 19 layers[25]. The VGG architecture is straightforward but highly effective because it just employs 3x3 convolutional filters. Medical image classification has made extensive use of VGG-19, particularly in applications where complicated pattern detection requires deep learning models.

Harsh Shah et al[26] presented an explicable approach for ascertaining cataracts using the VGG-19 convolutional neural network (CNN) and the Gradient-weighted Class Activation Mapping (Grad-CAM) visualization technique. They trained and tested model

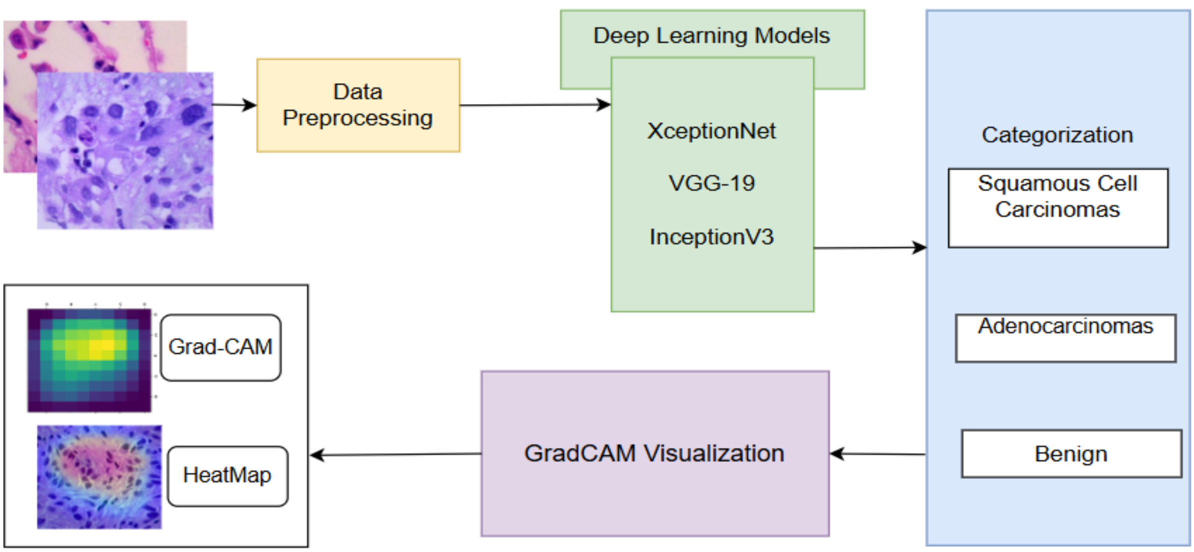


Fig. 2. Proposed Methodology

Manuscript received January 16, 2025; Accepted April 20, 2025; date of publication May 5, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.690

using 2,112 high-resolution fundus images from a publicly available dataset. According to the findings, the proposed method attained a very high accuracy of 97%. Furthermore, the Grad-CAM visuals showed the image regions that accord to the model's decision, providing discernment into the diagnosis process and enhancing the model's trustworthiness. The findings of this study could aid in ameliorating patient outcomes and lowering healthcare costs by allowing for the early recognition and diagnosis of cataracts. Overall, the fusion of VGG-19 with Grad-CAM offers a viable option in the medical domain for recognizing cataracts and understanding CNN - based decisions. Cleverson Marques Vieira et al[27] Machine learning models have spread throughout numerous sectors, revolutionizing illness detection and providing surprising applications in healthcare. In particular, the introduction of artificial intelligence approaches has dramatically altered the field of ophthalmology, allowing for the early diagnosis of neurodegenerative eye illnesses such as glaucoma via picture categorization. However, the lack of explainability in model judgements is a significant hurdle to their broad use in clinical practice. This study overcomes this issue by investigating and implementing explainable artificial intelligence (XAI) approaches on various convolutional neural network (CNN) architectures for glaucoma classification. The research is on providing ophthalmologists with reliable resources for human interpretation and clinical diagnosis. The work is based on a unique visual interpretation strategy known as SCIM (SHAP-CAM Interpretable Mapping) and compares its performance to current approaches.

III. Methods

The proposed model shown in Fig. 2 takes the resized images to a uniform size 299*299, normalize pixel values and apply data augmentation to enhance model generalization. The input is lung cancer dataset comprising of 15,000 histopathological images belonging to three categories namely lung adenocarcinomas, lung squamous cell carcinoma, and lung benign cells. The model is trained on InceptionV3, XceptionNet, VGG19 architectures. In order to enhance the interpretability of the predictions made by these deep learning models, the Grad-CAM (Gradient-weighted Class Activation Mapping) technique is integrated into the system. Grad-CAM operates by computing the gradients of the target class output with respect to the feature maps of the final convolutional layers. These gradients are then used to assign importance weights to the feature maps, allowing for the generation of heatmaps that highlight the most influential regions within the input images. These heatmaps provide critical visual explanations by indicating which parts of the histopathological tissue the model focuses on when making a diagnosis. The performance of InceptionV3, XceptionNet, and VGG19 are compared based on evaluation metrics and Grad-CAM is employed for interpretability.

A. Inceptionv3

InceptionV3, a deep convolutional neural network, is highly effective for lung cancer categorization from histopathological images due to its systematic design and multi-scale attribute extrication capabilities[28] . The model represented in Fig. 3 accepts input images of size 299x299x3, followed by several convolutional layers that extricate low-level attributes such as edges and textures. Its core strength lies in the Inception modules, which perform parallel convolutions with various kernel sizes (1x1, 3x3, 5x5) and pooling operations, allowing the model to apprehend both fine-grained details and broader patterns[29]. Factorized convolutions and asymmetric convolutions minimizes computational complexity while maintaining rich attribute representation. Auxiliary classifiers provide additional supervision during training, improving convergence. As the network progresses, reduction layers downsample attribute maps, retaining significant details. The final stages use Global Average Pooling (GAP) to minimize dimensionality before a fully

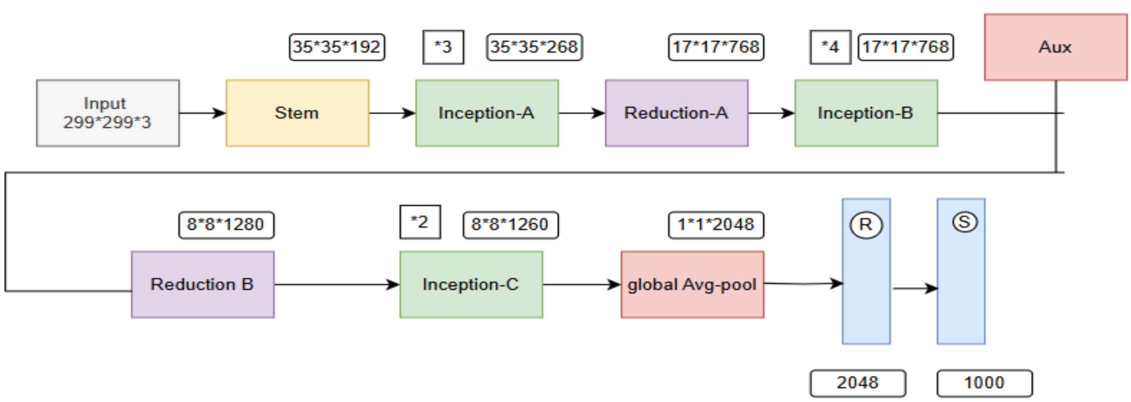


Fig. 3. InceptionV3 architecture

Manuscript received January 16, 2025; Accepted April 20, 2025; date of publication May 5, 2025
Digital Object Identifier (DOI): https://doi.org/10.35882/jeeemi.v7i3.690

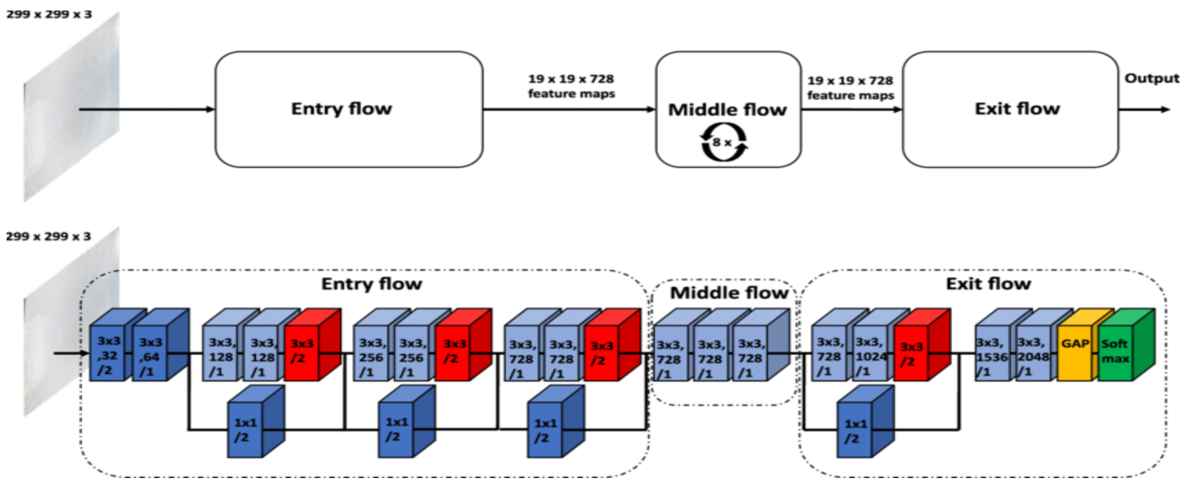


Fig. 4. XceptionNet architecture

connected layer and a softmax activation for final classification [30]. The model can be fine-tuned using pre-trained weights and offers improved comprehensibility through Grad-CAM, which generates heatmaps to envisage important regions in the lung tissue, helping clinicians understand the model's decision-making process. Overall, InceptionV3 is highly efficient for lung cancer detection, offering precise prognosis with enhanced interpretability and computational efficiency. The fundamental operation in InceptionV3 is the convolution, defined as in Eq. (1) [28].

$$y(i,j,k) = \sum m \sum n \sum c \ W(m.n,c,k).x(i-m,j-n,c) \qquad (1)$$

where m, n indices over the height and width of the kernel, c sums over input channels, k represents the index for output feature map, i and j depicts spatial position of output feature map, $x(i-m,j-n,c)$ represents input tensor for channel. $W(m.n,c,k)$ represents weight tensor filter and $y(i,j,k)$ is the output tensor for value at position (i,j) for the output feature m (channel k). Each inception module applies different sized filters in parallel as in Eq. (2)[28].

$$F_{out} = \oplus(f_1(x), f_3(x), f_5(x), f_{pool}(x) \qquad (2)$$

where$f_1(x)$is a 1 * 1 convolution for dimensionality reduction, $f_3(x)$ is a 3*3 convoltion for local attribute extrication, $f_5(x)$ is a 5*5 convolution for larger receptive fields,$f_{pool}(x)$ is a pooling operation for spatial compression ⊕ represents concatenation along the channel dimension. Inception V3 uses auxiliary classifiers at intermediate layers to mitigate vanishing gradients. The auxiliary loss is calculated as in Eq. (3)[28].

$$L_{aux} = \sum_i y_i \log y_i^{\wedge} \qquad (3)$$

where $y_i$is the ground truth label, $y_i^{\wedge}$ is the predicted probability. InceptionV3 employs batch normalization after every convolutional layer to stabilize training as defined in Eq. (4)[28].

$$x^{\wedge} = \frac{x-\mu_B}{\sqrt{\sigma_B^2+\epsilon}} \qquad (4)$$

μ_B and σ_B^2 are the batch mean and variance. ϵ is a small constant to avoid division by zero.

B. XCEPTIONNET

XceptionNet, an extension of the Inception architecture, is a powerful deep learning model well suited for lung cancer maps. In XceptionNet shown in Fig. 4, each standard convolution operation is replaced with two classifications from histopathological images [31]. The architecture is based on depthwise separable convolutions which significantly improve the computational efficiency by separating the process of filtering and combining features smaller operations: a depthwise convolution (incorporating a single filter to each input channel) and a pointwise convolution (1x1 convolutions that amalgamates the output from the depthwise convolution) cancer tissue[32]. XceptionNet replaces standard convolutions with depthwise separable convolutions, which are decomposed into:

1. Depthwise Convolution

Each channel of the input is convolved separately with a different filter as in Eq. (5)[23].

$$Y_d^c = W_d^c * X_c \qquad (5)$$

whereY_d^c is the depthwise convolved output for channel c, W_d^c is the depthwise convolution filter for channel c,X_c is the input channel c.

2. Pointwise Convolution

Manuscript received January 16, 2025; Accepted April 20, 2025; date of publication May 5, 2025
Digital Object Identifier (DOI): https://doi.org/10.35882/jeeemi.v7i3.690

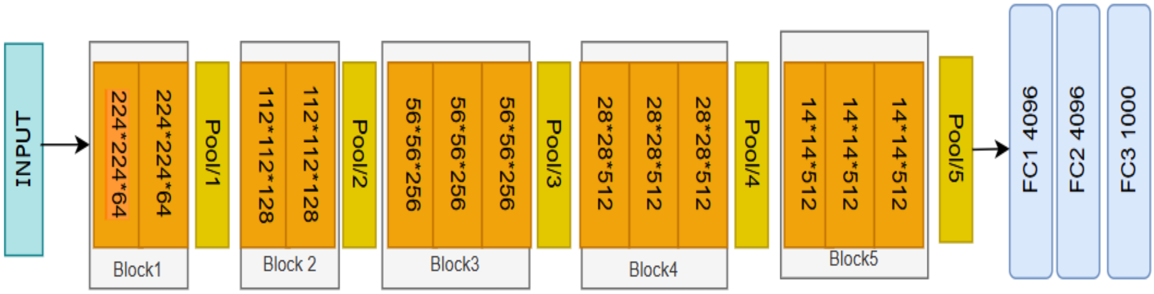


Fig. 5.VGG-19 architecture

A 11×1 convolution is applied across all depthwise-filtered channels to mix channel information as depicted in Eq. (6)[23].

$$Y_d^c = \sum c\, W_d^c * X_c \tag{6}$$

where Y_d^c is the final output feature map, $W_d^c$is the pointwise convolution filter for channel c.

3. Xception Block
Each Xception block consists of:
a) Depthwise separable convolutions followed by Batch Normalization and ReLU activation
b) Skip connections (Residual connections) to enhance gradient flow.

The Eq. (7) [33] represents the single Xception block

$$F(X) = \sigma(BN(W_p * \left(\sigma\left(BN(W_d * X)\right)\right)) \tag{7}$$

where W_p and W_d are depthwise and pointwise convolution filters, BN is batch normalization and σ is the ReLU activation function.

4. Residual Connection
XceptionNet uses residual connections, which add the input to the output of the convolution block as depicted in Eq. (8)[33].

$$\mathrm{Y} = \mathrm{X} + \mathrm{F(X)} \tag{8}$$

where X is a input to block , F(X) is transformation applied to block and Y is the output of residual connection. If dimensions do not match, a projection layer Ws is applied as in Eq. (9)[33].

$$\mathrm{Y} = \mathrm{W_s} * \mathrm{X} + \mathrm{F(X)} \tag{9}$$

where Ws is the learned linear projection applied to X.

C. VGG-19

VGG19 is a deep convolutional neural network known for its simplicity and depth, which makes it effective for image classification tasks, including lung cancer detection from histopathological images. The architecture depicted in Fig. 5 consists of 19 layers, with 16 convolutional layers and 3 fully connected layers. VGG19 uses small 3x3 convolution filters, which are stacked to capture complex patterns while maintaining a manageable number of parameters[34]. The network also employ max-pooling layers after every few convolutional layers to reduce spatial dimensions and retain essential features. After the convolutional layers, flattening is applied to reshape the output into a 1D vector, which is then passed through fully connected layers. The penultimate layer uses softmax activation to output probabilities for classification, distinguishing between cancerous and non-cancerous lung tissue. Although VGG19 is computationally more expensive compared to models like InceptionNet or XceptionNet, its straightforward design and effectiveness in capturing detailed spatial information make it a valuable tool for lung cancer classification when fine-tuned on a large histopathological dataset. Table 2 represents the configuration settings for the various architectures in the proposed model with respect to input shape, learning rate, batch size, optimizer, loss function and drop out parameters.

Table 2. Configuration Settings

Parameter	InceptionV3	Xception Net	VGG-19
Input shape	299*299*3	299*299*3	224*224*3
Learning Rate	0.0001	0.0001	0.0001
Batch Size	32	32	32
Optimizer	Adam	SGD	RMSProp
Loss Function	Categorical cross entropy	Categorical cross entropy	Categorical cross entropy
DropOut	0.4	0.3	NIL

D. Grad- CAM

Step 1: Gradient Computation in GradCAM
In order to calculate the final score for a particular class, each attribute map from the penultimate convolutional layer records various amounts of high-

Manuscript received January 16, 2025; Accepted April 20, 2025; date of publication May 5, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.690

level information about the input image. The goal is to investigate the connection between the output and the attribute maps. A modification to the attribute map would alter the score value for any particular class, c . In order to highlight the areas of the input image that were used to produce the prediction, importance score must be calculated based on the gradient of the class score z^c with respect to the attribute maps M^k as represented in Eq. (10)[12].

$$\frac{\partial z^c}{\partial M^k} \quad (10)$$

For the equation above, Z^c, is a scalar (anticipated score before the SoftMax calculation), and M^k is a two-dimensional attribute map. So, the gradient is also a two-dimensional map with the exact spatial dimensions as the attribute maps, M^k.

Step 2: Enumerate Alpha Values for GradCAM

Alpha values, which can be regarded as significant values, are obtained in this stage by performing global average pooling of the gradients throughout the width and height as represented in Eq. (11)[12].

$$\alpha_k^c = \frac{1}{Y}\sum_i\sum_j \frac{\partial y^c}{\partial A_{ij}^k} \quad (11)$$

where Y is the sum of number of pixels in each attribute map, and the calculation below yields k alpha values for each class.

Step 3: Produce GradCAM Heatmap

The alpha values determined in the previous step are analogous to the weights in the penultimate layer for the CAM approach, excluding that the alpha values were generated using gradients.

Similar to CAM, a weighted sum of the activation maps can be calculated where the weights are the alpha values determined as below in Eq. (12) [34].

$$\alpha_k^c M^k = \alpha_1^c . M^1 + \alpha_2^c . M^2 \dots + \alpha_k^c . M^k \quad (12)$$

where α_k^c is importance weight for feature map k for class c, c is the specific class and k is total number of features.

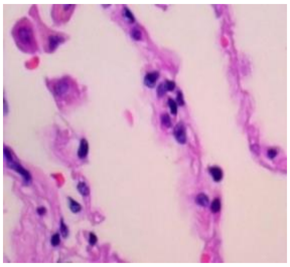
Step 4: Eventually, to determine the GradCAM heatmap, the above-aggregated sum is passed through a ReLU activation function to zero out any negative gradients as in Eq. (13) [34].

$$L_{Grad-CAM}^c = SiLU\ (\textstyle\sum_k \alpha_k^c M^k \quad (13)$$

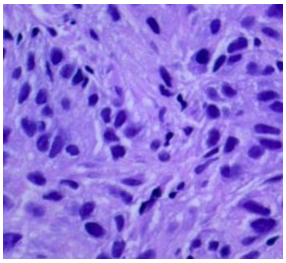
where $L_{Grad-CAM}^c$ is the Grad-CAM heatmap for class c and SiLU is the Sigmoid Linear Unit activation function. The aforementioned computation will yield a coarse localization map with precisely the same spatial dimensions as the attribute maps. The heatmap is then upsampled to the same size as the input image and normalized to the [0,1] range to create a final heatmap that can be superimposed over the image.

IV. Results

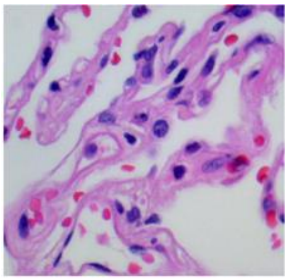
The Larxel's lung cancer dataset from Kaggle is used for experimental purpose. There are 15,000 histopathological scans in this dataset, separated into three categories of lung adenocarcinomas, lung squamous cell carcinomas, and lung benign tissues. The images were obtained using a sample of 750 original lung tissue images from sources that complied with and were authorized by the Health Insurance Portability and Account ability Act (250 benign lung tissue, 250 lung adenocarcinomas, and 250 lung squamous cell carcinomas). To expand the dataset and enhance the diversity of training samples, extensive data augmentation was performed using the Augmentor package. Through various augmentation techniques such as rotations, flips, scaling, and color adjustmentsthe original 750 images were artificially increased to form a total of 15,000 images.



(a)



(b)



(c)

Fig. 6. Misclassified images by InceptionV3, XceptionNet, VGG-19 respectively a) Sample normal image misclassified as adenocarcinomas, (b) Sample adenocarcinomas misclassified as squamous cell carcinomas, (c) Sample normal image misclassified as adenocarcinomas.

This augmentation process ensured that each of the three categories was equally represented with 5,000 images, which helps in improving the robustness and generalization capability of the models trained on this dataset. The split ratio for the training, validation, and testing sets is 70:10:20. 10500 of the 15,000 images are used for training, 1500 for validation and 3000 for testing. Each model is trained using a training set (70%), validation set (10%), Test Set (20%).

Manuscript received January 16, 2025; Accepted April 20, 2025; date of publication May 5, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.690

Categorical cross-entropy for multi-class with Adam optimizer is applied. The training is explicitly set for 90 epochs with 328 steps per epoch and 93 steps for validation.The optimizer is defined as the SGD optimizer with a learning rate of 1e^{-3} for XceptionNet, Adam optimizer with a learning rate of 1e^{-4} for InveptionV3,the Adam optimizer with learning rate of 1e^{-4}, is used for VGG-19 which is well-suited for fine-tuning deep learning models with categorical_crossentropy as the loss function.

Fig. 6 illustrates examples of misclassified images from three deep learning models—InceptionV3, XceptionNet, and VGG-19—using Grad-CAM visualizations to highlight regions that influenced the models' predictions. In (a), a normal lung tissue image was incorrectly classified as adenocarcinoma, suggesting that the model may have focused on benign structural patterns resembling malignant features. In (b), an actual adenocarcinoma sample was misclassified as squamous cell carcinoma, indicating confusion between similar histopathological patterns. Lastly, in (c), another normal image was again misclassified as adenocarcinoma, emphasizing the challenge in distinguishing subtle variations in tissue appearance. These Grad-CAM visualizations provide insight into the decision-making processes of the models and reveal potential areas where the models may be overfitting or misinterpreting key features.

A. Evaluation Metrics and Results

(i) Accuracy is one of the most commonly used performance metrics for evaluating the overall effectiveness of a classification model. It represents the ratio of correctly predicted samples (both true positives and true negatives) to the total number of samples in the dataset. In other words, accuracy shows how often the model makes correct predictions. Although accuracy provides a simple and intuitive measure, it can be misleading when dealing with imbalanced datasets. For example, if the dataset has a dominant class, the model may achieve high accuracy by predicting the majority class for all samples, while performing poorly on the minority class. As such, accuracy should be used in conjunction with other metrics like precision, recall, and F1 score to get a more comprehensive evaluation of model performance. Eq. (14)[31] represents the accuracy.

$$\text{Accuracy}=\frac{True\ Positive+True\ Negative}{Total\ Images\ Sampled} \tag{14}$$

(ii) Precision: Precision, also known as positive predictive value, is a metric that focuses on the performance of the model in predicting the positive class. Specifically, precision is the ratio of true positive predictions to the total number of samples predicted as positive (true positives + false positives). This metric is particularly important in scenarios where false positives are costly or undesirable. For instance, in medical diagnosis, a high precision means that when the model predicts a patient has a disease, it is very likely to be correct. It is the ratio of truly prognosed positive samples to overall prognosed positive samples as in Eq. (15)[31].

$$\text{Precision} = \frac{True\ Positive}{True\ Positive+False\ Positive} \tag{15}$$

(iii) Recall: Recall, also known as sensitivity or true positive rate, measures how well the model identifies actual positive samples. It is the ratio of true positives to the total number of actual positive samples in the dataset (true positives + false negatives). Recall is critical in situations where missing a positive instance is more detrimental than falsely classifying a negative instance as positive. For example, in disease detection, recall would measure how effectively the model detects all the patients who actually have the disease. A high recall means fewer cases of the condition are overlooked, but it may result in more false positives. To obtain a balanced evaluation of model performance, recall is typically considered alongside precision. It is the ratio of truly prognosed positive samples to overall samples in the actual class as in Eq.(16)[31].

$$\text{Recall} =\frac{True\ Positive}{True\ Positive+Fasle\ Negative} \tag{16}$$

(iv) F1 score: The F1 score is a harmonic mean of precision and recall, providing a single metric that balances the trade-off between the two. It is especially useful when the class distribution is imbalanced and you need to balance the importance of both precision and recall. The F1 score is calculated as the ratio of twice the product of precision and recall to the sum of precision and recall. This metric is valuable because it ensures that a model is not biased towards only optimizing one metric (such as precision) at the expense of the other (such as recall). A high F1 score indicates that the model performs well in both identifying positive instances and minimizing false positives. In cases where a balance between false positives and false negatives is critical, the F1 score is a preferred choice. It is the weighted mean of precision and recall as in Eq. (17)[31].

$$\text{F1 score} = \frac{2*(Precision*Recall)}{Precision+Recall} \tag{17}$$

The performance of the three modelsInceptionNetV3, XceptionNet, and VGG-19 was evaluated using several key metrics: accuracy, precision, F1-score, and recall. XceptionNet outperforms the other two models across all metrics, with an accuracy of 98.2%, indicating the

Manuscript received January 16, 2025; Accepted April 20, 2025; date of publication May 5, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.690

Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 7, No. 3, July 2025, pp: 567-580 e-ISSN: 2656-8632

highest overall classification performance. It also leads in precision at 97.4%, meaning it has the highest proportion of correct positive predictions, minimizing false positives.

Table 3. **Evaluation Metrics of the Proposed Model**

Models	Acc.	Precision	F1-Score	Recall
Inception V3	97.1	96.8	97.2	96.4
Xception Net	98.2	97.4	98.6	97.7
VGG-19	95.8	96.4	95.3	94.8

Additionally, XceptionNet achieves the highest F1-score of 98.6%, which highlights its well-balanced performance between precision and recall. Recall for XceptionNet is 97.7%, indicating that it correctly identifies a large portion of actual positive cases. In comparison, InceptionNetV3 shows slightly lower performance across all metrics, with an accuracy of 97.1%, precision of 96.8%, F1-score of 97.2%, and recall of 96.4%. VGG-19 has the lowest performance, with an accuracy of 95.8%, precision of 96.4%, F1-score of 95.3%, and recall of 94.8%, indicating it struggles slightly more in both precision and recall, especially in identifying all positive cases. Overall, XceptionNet provides the best balance and highest performance across these evaluation metrics. The comparative analysis of three CNN models on Categorizing the lung cancers are shown in Fig. 7 for VGG19, InceptionV3 and XceptionNet models.

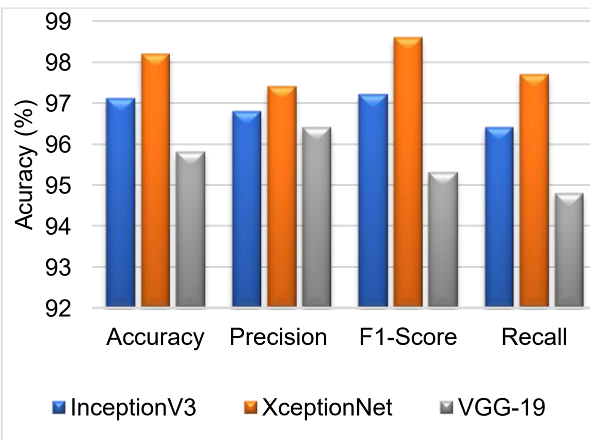


Fig. 7. **Comparative Analysis of CNN Models**

V. DISCUSSION

The key findings from the analysis of misclassified samples using Grad-CAM and Heatmaps for InceptionV3, XceptionNet, and VGG-19 reveals that each model has distinct weaknesses in feature focus, which contribute to misclassifications. InceptionV3 tends to overemphasize complex or irrelevant patterns in the image, indicating that the model may be overfitting to specific features, which do not generalize well. XceptionNet, known for its depth wise separable convolutions, often concentrates on small, localized regions, which may include textures or artifacts not relevant to the target class, leading to misclassifications. VGG-19, with its simpler architecture, struggles with capturing global context and often focuses on smaller, less informative regions, causing the model to miss important contextual cues. These findings highlight the importance of improving model attention mechanisms and feature extraction capabilities to focus more effectively on the relevant areas of the image. This analysis emphasizes the need for better generalization and context understanding in these models to reduce misclassification errors. The heatmaps and GradCAM for misclassified samples are shown in Fig. 8. Furthermore, these insights suggest that improving model architectures to better capture both global and local features, as well as regularizing the models to avoid overfitting, could enhance their performance. Incorporating more advanced attention mechanisms or multi-scale features may also help models focus on the most relevant parts of the image, reducing misclassifications.

Table 4. **Comparison with state of art models**

Authors	Methods	Accuracy (%)
Sabbir Ahmed et al [35]	Logistic Regression , SHAP and LIME	97%
Joshua et al [36]	SENET (Squeeze-and-Excitation Networks), GradCAM++	97.08
Niyaz Ahmad et al [37]	Convolutional Neural Network, XGBoost, SHapley Additive exPlanations	97.43
Proposed Model	**Xception Net, GradCAM**	**98.2**

The performance evaluation of the proposed model with existing state of art models is discussed in Table 4. Sabbir et al[35] have proposed an interpretable lung cancer diagnostic system utilizing various machine learning models, including Decision Tree, Logistic Regression, Random Forest, and Naive Bayes classifiers. Among the models evaluated, Logistic Regression and Random Forest achieved the highest prognosis accuracy of 97% for lung cancer recognition. To enhance the transparency and interpretability of our

Manuscript received January 16, 2025; Accepted April 20, 2025; date of publication May 5, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.690

models, we also employed two widely used explainable AI (XAI) techniques: SHAP and LIME.

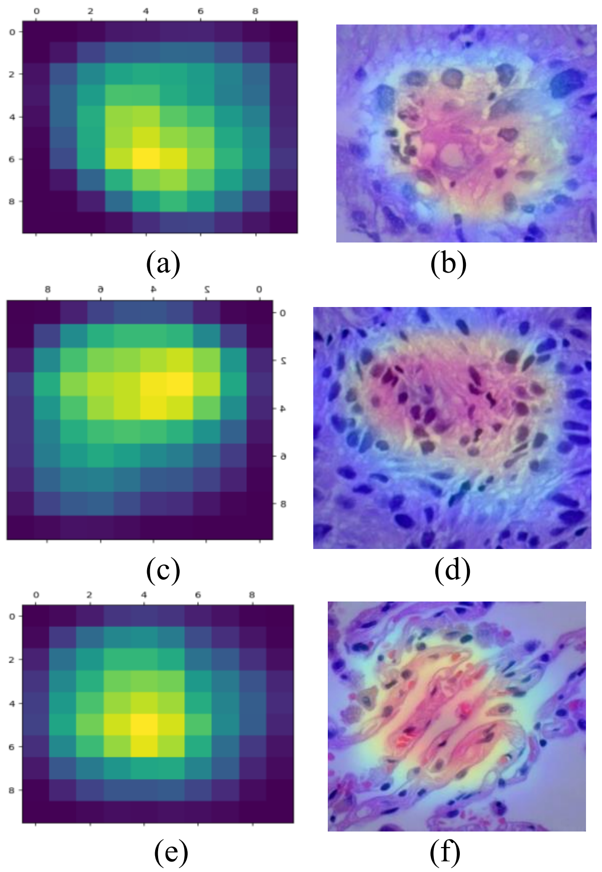


Fig. 8. HeatMaps and Grad-CAM for misclassified samples respectively, (a),(b) Misclassified sample by InceptionV3 (c),(d) Misclassified sample by XceptionNet, (e),(f) Misclassified sample by VGG-19.

Joshua et al[36] have introduced an advanced approach for distinguishing malignant from benign lung nodules in CT scans by integrating Grad-CAM++ with a Squeeze-and-Excitation Network (SENET). The proposed SENET-Grad-CAM++ module leverages the feature calibration and enhancement capabilities of SENET to significantly improve feature discriminability in lung cancer classification. Evaluated using the publicly available Lung dataset comprising 1,230 nodules (600 malignant and 630 benign) the method achieved an AUC of 0.9664 and an accuracy of 97.08%. These promising results highlight the robustness of the proposed technique in nodule classification and underscore its potential to support radiologists in accurately interpreting diagnostic CT scans and distinguishing between benign and malignant lung nodules.

Niyaz ahmad et al[37] have presented DeepXplainer, a novel interpretable hybrid deep learning framework for lung cancer detection and prediction explainability. The model integrates a Convolutional Neural Network (CNN) for automatic feature extraction with XGBoost for final class label prediction. To enhance transparency, the SHAP (SHapley Additive exPlanations) method is applied to interpret predictions at both local and global levels. The approach has demonstrated superior performance across multiple metrics. It achieved an accuracy of 97.43%, sensitivity of 98.71%, and an F1-score of 98.08%, outperforming existing methods while offering clear interpretability of each prediction. This interpretability is crucial for medical practitioners, enabling them to confidently rely on the model's predictions for lung cancer detection and diagnosis.

The proposed model, XceptionNet with Grad-CAM, achieved an accuracy of 98.2%, which demonstrates its strong performance in lung cancer detection. While InceptionV3 and VGG-19 may struggle with focusing on irrelevant features or context, XceptionNet utilizes depthwise separable convolutions that allow it to focus more effectively on relevant patterns in the image. The integration of Grad-CAM also enhances model interpretability, providing clear visual explanations for its predictions, which is critical in medical applications. This makes XceptionNet a robust and interpretable choice for lung cancer detection, surpassing traditional models in both performance and transparency.

The limitations of the models evaluated, such as InceptionV3, XceptionNet, and VGG19, are fine-tuned for lung cancer histopathological images, but their performance may be suboptimal when applied to other medical imaging modalities, such as CT scans or MRIs. While Grad-CAM is useful, combining it with other interpretability methods, such as Layer-wise Relevance Propagation (LRP) or Integrated Gradients, could provide a more comprehensive understanding of model decision-making.

The implications of this study are significant, particularly in the context of clinical decision support for lung cancer diagnosis. By integrating Grad-CAM with deep learning models like InceptionV3, XceptionNet, and VGG19, this study not only enhances the accuracy of automated lung cancer detection from histopathological images but also provides much-needed interpretability. The ability to visualize which areas of the tissue samples influence a model's decision is crucial for building trust in AI-based medical systems. This transparency helps clinicians understand the rationale behind predictions, offering a second opinion and reducing diagnostic errors. Moreover, the insights gained can assist in identifying critical features that are indicative of cancerous tissues, enabling more precise diagnoses and potentially improving patient outcomes. The study also lays the

Manuscript received January 16, 2025; Accepted April 20, 2025; date of publication May 5, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.690

groundwork for future applications of AI in pathology, where interpretability is essential for clinical acceptance and integration into routine practice.

VI. CONCLUSION

This study aims to enhance the transparency and interpretability of deep learning models used in the classification of lung cancer histopathological images by employing the Grad-CAM (Gradient-weighted Class Activation Mapping) technique. The core objective is to provide meaningful visual explanations that identify the most critical regions in tissue images influencing the model's predictions, thereby ensuring that the decision-making process is understandable and aligned with clinical reasoning. Grad-CAM was systematically applied to several prominent convolutional neural network architectures, including InceptionV3, XceptionNet, and VGG19, each known for distinct architectural strengths such as multi-scale feature detection, deep pattern abstraction, and global feature extraction. The visualizations produced by Grad-CAM effectively demonstrated where each model focused during classification, offering vital insights into the interpretability of AI-driven diagnosis. Among the evaluated models, XceptionNet exhibited superior classification performance, achieving the highest accuracy and generating clearer, more clinically relevant heatmaps that highlighted abnormal and cancerous regions with high precision. These results not only reinforced the reliability of the model predictions but also provided pathologists with supportive diagnostic evidence through visual interpretation.

The study evaluated the performance of three deep learning models for lung cancer detection: InceptionV3, XceptionNet, and VGG-19. XceptionNet achieved the highest accuracy at 98.2%, with a corresponding error rate of 1.8%. InceptionV3 followed closely with an accuracy of 97.1%, resulting in an error rate of 2.9%. VGG-19, although performing well, had the lowest accuracy among the three models, reaching 95.8%, which corresponds to an error rate of 4.2%. These results demonstrate that while all models perform admirably, XceptionNet outperforms the others in terms of accuracy and minimizes the error rate, making it a more reliable choice for lung cancer detection in histopathological images. The findings suggest that integrating Grad-CAM into the diagnostic workflow can bridge the gap between complex AI models and clinical practice, fostering greater trust among healthcare professionals. Additionally, comparing the heatmaps across different architectures revealed complementary strengths, suggesting the potential for ensemble approaches to further enhance diagnostic performance. Grad-CAM provides qualitative heatmaps, but it lacks a standardized quantitative measure to evaluate the reliability of highlighted regions. For future work, techniques like Guided Grad-CAM and SmoothGrad can refine heatmaps by reducing noise and enhancing fine-grained details, helping in more precise localization of cancerous regions and improving spatial accuracy. Instead of relying solely on the final convolutional layers, analyzing heatmaps from multiple layers can provide a hierarchical view of feature extraction, improving interpretability.. Finally, collaborative studies involving radiologists and pathologists are planned to validate the clinical usefulness of the generated visual interpretations and refine the system for real-world deployment.

REFERENCES

[1] W. Saeed and C. Omlin, “Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities,” *Knowl Based Syst*, 2023, doi: 10.1016/j.knosys.2023.110273.
[2] S. S Band *et al.*, “Application of explainable artificial intelligence in medical health: A systematic review of interpretability methods,” *Inform Med Unlocked*, 2023, doi: 10.1016/j.imu.2023.101286.
[3] E. Tjoa and C. Guan, “A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI,” *IEEE Trans Neural Netw Learn Syst*, 2021, doi: 10.1109/TNNLS.2020.3027314.
[4] I. D. Mienye and Y. Sun, “A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects,” 2022. doi: 10.1109/ACCESS.2022.3207287.
[5] A. Singh, S. Sengupta, and V. Lakshminarayanan, “Explainable deep learning models in medical image analysis,” 2020. doi: 10.3390/JIMAGING6060052.
[6] M. A. Ganaie, M. Hu, A. K. Malik, M. Tanveer, and P. N. Suganthan, “Ensemble deep learning: A review,” 2022. doi: 10.1016/j.engappai.2022.105151.
[7] K. Gao, H. Shen, Y. Liu, L. Zeng, and D. Hu, “Dense-CAM: Visualize the Gender of Brains with MRI Images,” in *Proceedings of the International Joint Conference on Neural Networks*, 2019. doi: 10.1109/IJCNN.2019.8852260.
[8] A. Teramoto, T. Tsukamoto, Y. Kiriyama, and H. Fujita, “Automated Classification of Lung Cancer Types from Cytological Images Using Deep Convolutional Neural Networks,” *Biomed Res Int*, 2017, doi: 10.1155/2017/4067832.
[9] R. Ibrahim and M. Omair Shafiq, “Explainable Convolutional Neural Networks: A Taxonomy, Review, and Future Directions,” 2023. doi: 10.1145/3563691.
[10] T. Patil and S. Arora, “Survey of Explainable AI Techniques: A Case Study of Healthcare,” in

Manuscript received January 16, 2025; Accepted April 20, 2025; date of publication May 5, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.690

Lecture Notes in Networks and Systems, 2023. doi: 10.1007/978-981-99-5652-4_30.

[11] M. Xiao, L. Zhang, W. Shi, J. Liu, W. He, and Z. Jiang, "A visualization method based on the Grad-CAM for medical image segmentation model," in *2021 International Conference on Electronic Information Engineering and Computer Science, EIECS 2021*, 2021. doi: 10.1109/EIECS53707.2021.9587953.

[12] 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*, 2017. doi: 10.1109/ICCV.2017.74.

[13] Z. Chen, Z. Tian, J. Zhu, C. Li, and S. Du, "C-CAM: Causal CAM for Weakly Supervised Semantic Segmentation on Medical Image," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2022. doi: 10.1109/CVPR52688.2022.01138.

[14] G. Visani, E. Bagli, and F. Chesani, "OptiLIME: Optimized lime explanations for diagnostic computer algorithms," in *CEUR Workshop Proceedings*, 2020.

[15] M. T. Ribeiro, S. Singh, and C. Guestrin, "'why should i trust you?' explaining the predictions of any classifier," in *NAACL-HLT 2016 - 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Demonstrations Session*, 2016. doi: 10.18653/v1/n16-3020.

[16] D. E. B. Schiavon, C. D. L. Becker, V. R. Botelho, and T. A. Pianoski, "Interpreting Convolutional Neural Networks for Brain Tumor Classification: An Explainable Artificial Intelligence Approach," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2023. doi: 10.1007/978-3-031-45389-2_6.

[17] M. M. Ahsan, R. Nazim, Z. Siddique, and P. Huebner, "Detection of covid-19 patients from ct scan and chest x-ray data using modified mobilenetv2 and lime," *Healthcare (Switzerland)*, 2021, doi: 10.3390/healthcare9091099.

[18] C. Patrício, J. C. Neves, and L. F. Teixeira, "Explainable Deep Learning Methods in Medical Image Classification: A Survey," *ACM Comput Surv*, 2023, doi: 10.1145/3625287.

[19] X. Kong, S. Liu, and L. Zhu, "Toward Human-centered XAI in Practice: A survey," 2024. doi: 10.1007/s11633-022-1407-3.

[20] M. A. S. Al Husaini, M. H. Habaebi, T. S. Gunawan, M. R. Islam, E. A. A. Elsheikh, and F. M. Suliman, "Thermal-based early breast cancer detection using inception V3, inception V4 and modified inception MV4," *Neural Comput Appl*, 2022, doi: 10.1007/s00521-021-06372-1.

[21] S. Bharati, M. R. H. Mondal, and P. Podder, "A Review on Explainable Artificial Intelligence for Healthcare: Why, How, and When?," *IEEE Transactions on Artificial Intelligence*, 2024, doi: 10.1109/TAI.2023.3266418.

[22] P. Rajpurkar and M. P. Lungren, "The Current and Future State of AI Interpretation of Medical Images," *New England Journal of Medicine*, 2023, doi: 10.1056/nejmra2301725.

[23] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017. doi: 10.1109/CVPR.2017.195.

[24] B. U. Maheswari *et al.*, "Explainable deep-neural-network supported scheme for tuberculosis detection from chest radiographs," *BMC Med Imaging*, 2024, doi: 10.1186/s12880-024-01202-x.

[25] M. Melinda, M. Oktiana, Y. Nurdin, I. Pujiati, M. Irhamsyah, and N. Basir, "Performance of ShuffleNet and VGG-19 Architectural Classification Models for Face Recognition in Autistic Children," *Int J Adv Sci Eng Inf Technol*, 2023, doi: 10.18517/ijaseit.13.2.18274.

[26] H. Shah, R. Patel, S. Hegde, and H. Dalvi, "XAI Meets Ophthalmology: An Explainable Approach to Cataract Detection Using VGG-19 and Grad-CAM," in *2023 IEEE Pune Section International Conference, PuneCon 2023*, 2023. doi: 10.1109/PuneCon58714.2023.10450053.

[27] C. M. Vieira, M. V. D. C. Oliveira, M. D. P. Guimarães, L. Rocha, and D. R. C. Dias, "Applied Explainable Artificial Intelligence (XAI) in the classification of retinal images for support in the diagnosis of Glaucoma," in *ACM International Conference Proceeding Series*, 2023. doi: 10.1145/3617023.3617026.

[28] H. Jin and S. Chen, "Biometric Recognition Based on Recurrence Plot and InceptionV3 Model Using Eye Movements," *IEEE J Biomed Health Inform*, 2023, doi: 10.1109/JBHI.2023.3313261.

[29] P. Bedi, N. Ningshen, S. Rani, and P. Gole, "Explainable Predictions for Brain Tumor Diagnosis Using InceptionV3 CNN Architecture," in *Lecture Notes in Networks and Systems*, 2024. doi: 10.1007/978-981-99-4071-4_11.

[30] P. Theerthagiri and G. B. Nagaladinne, "Deepfake Face Detection Using Deep InceptionNet Learning Algorithm," in *2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science, SCEECS*

Manuscript received January 16, 2025; Accepted April 20, 2025; date of publication May 5, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.690

2023, 2023. doi: 10.1109/SCEECS57921.2023.10063128.
[31] R. Mothkur and B. N. Veerappa, “Deep Learning-Based Three Type Classifier Model for Non-small Cell Lung Cancer from Histopathological Images,” in *Lecture Notes in Networks and Systems*, 2023, pp. 481–493. doi: 10.1007/978-981-19-9379-4_35.
[32] M. Mateen, J. Wen, Nasrullah, S. Song, and Z. Huang, “Fundus image classification using VGG-19 architecture with PCA and SVD,” *Symmetry (Basel)*, vol. 11, no. 1, 2019, doi: 10.3390/sym11010001.
[33] X. Lu and Y. A. Firoozeh Abolhasani Zadeh, “Deep Learning-Based Classification for Melanoma Detection Using XceptionNet,” *J Healthc Eng*, 2022, doi: 10.1155/2022/2196096.
[34] H. Zhang and K. Ogasawara, “Grad-CAM-Based Explainable Artificial Intelligence Related to Medical Text Processing,” *Bioengineering*, 2023, doi: 10.3390/bioengineering10091070.
[35] M. S. Ahmed, K. N. Iqbal, and M. G. R. Alam, “Interpretable Lung Cancer Detection using Explainable AI Methods,” in *2023 International Conference for Advancement in Technology, ICONAT 2023*, 2023. doi: 10.1109/ICONAT57137.2023.10080480.
[36] E. S. N. Joshua, D. Bhattacharyya, M. Chakkravarthy, and H. J. Kim, “Lung cancer classification using squeeze and excitation convolutional neural networks with grad Cam++ class activation function,” *Traitement du Signal*, 2021, doi: 10.18280/ts.380421.
[37] N. A. Wani, R. Kumar, and J. Bedi, “DeepXplainer: An interpretable deep learning based approach for lung cancer detection using explainable artificial intelligence,” *Comput Methods Programs Biomed*, 2024, doi: 10.1016/j.cmpb.2023.107879.

AUTHORS BIOGRAPHY

Rashmi Mothkur working as Associate Professor in the Department of Alliance School of Advanced Computing at Alliance University, Bangalore. She has completed her Ph.D from Visweswaraya Technological University, Belgaum. Her research work is based on the medical image processing, deep learning and bioinspired techniques. She has completed her M.Tech degree from M.S.Ramiah Institute of Technology, Bangalore. She has completed her B.Tech degree from Bapuji Institute of Engineering and Technology, Davangere. She has total of 11 years of teaching experience. Her research interest include Artificial Intelligence, Medical Image Processing, Computer Vision. She has published various conference papers and journal papers.

PULLAGURA SOUBHAGYALAKSHMI working as Associate Professor in the Department of Computer Science & Engineering, K.S. Institute of Technology, Bangalore. She has completed her Ph. D from Visvesvaraya Technological University, Belgaum. She has secured first class with distinction in M.Tech, Computer Science and Engineering from Jawaharlal Nehru Technological University, Hyderabad in 2006. She has received her B.Tech in Computer Science Engineering from G Pulla Reddy Engineering College in 1999, Kurnool and Andra Pradesh.She has 12 years and 9 months of teaching experience. Her research interest include Computer Networks, Security, Image Processing, and Artificial Intelligence. She has published various conference papers and journal papers.

Swetha C. B is currently serving as an Assistant Professor at the Alliance School of Advanced Computing, Alliance University, Bengaluru. She is pursuing her Ph.D. in the field of Federated Learning. She earned her B. Tech degree in Computer Science and Engineering from MG University, Kerala. Additionally, she holds an M. Tech in Information Systems Security from IGNOU, completed in a regular face-to-face mode. With over seven years of teaching experience, she has been actively engaged in academia, mentoring students, and contributing to research. Her research interests include Artificial Intelligence and Cybersecurity. She has enriched her knowledge and skills by actively participating in numerous workshops, seminars, and faculty development programs.

Manuscript received January 16, 2025; Accepted April 20, 2025; date of publication May 5, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.690