

Received 25 June 2025, accepted 9 July 2025, date of publication 14 July 2025, date of current version 18 July 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3588407

## RESEARCH ARTICLE

# Multi-Modal Deep Learning for Lung Cancer Detection Using Attention-Based Inception-ResNet

MOHAMED HOSNY<sup>ID</sup>, IBRAHIM A. ELGENDY<sup>ID</sup>, AND MOUSA AHMAD ALBASHRAWI

IRC for Finance and Digital Economy, KFUPM Business School, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia

Corresponding author: Mohamed Hosny (mohamed.elnogomy@kfupm.edu.sa)

This work was supported by the Interdisciplinary Research Center for Finance and Digital Economy (IRC-FDE), King Fahd University of Petroleum and Minerals (KFUPM), Saudi Arabia under Project INFE2414.

**ABSTRACT** Lung cancer is one of the deadliest malignancies worldwide, demanding swift and accurate diagnosis for effective treatment. Traditional screening methods rely on manual interpretation of medical images. However, these methods are time-intensive and highly susceptible to human error. Deep learning (DL) has emerged as a powerful alternative to autonomously identify complex patterns within radiological and histopathological images. Nevertheless, existing DL models for lung cancer detection suffer from critical limitations, including reliance on single imaging modalities, insufficient datasets, suboptimal feature extraction techniques and constrained generalizability. Accordingly, we introduce a novel DL framework that blends diverse imaging modalities, including X-rays, computed tomography (CT) and histopathological images. The proposed framework employs Inception-ResNet module to extract multi-scale spatial features and refine deep feature representations through residual learning. This hybrid module combines the convolutional pathways of Inception architectures with the gradient optimization benefits of residual connections. Besides, the proposed architecture is embedded with sequential multi-scale convolutional fusion and efficient channel attention mechanisms to ameliorate feature diversity and optimize feature importance. These components aid the model to focus on highly discriminative regions within medical images. The proposed model attained an accuracy of 95.35%, 99.68%, 99.73% and 99.26% using X-ray, CT, histopathological and mega datasets, respectively. Comparative experiments unveiled that the proposed model outperformed conventional DL architectures in lung cancer detection. The proposed system, utilizing advanced attention mechanisms and multi-modal imaging capabilities, has the potential to revolutionize early lung cancer diagnosis and extend its impact to other critical diseases. This work represents a paradigm shift in medical image analysis through bridging the gap between DL and clinical applications. The proposed model is available at <https://github.com/MohamedHosny90/Lung-Cancer-Detection.git>

**INDEX TERMS** Multi-modal imaging, lung cancer detection, hierarchical feature extraction, efficient channel attention.

## I. INTRODUCTION

Lung cancer remains one of the most formidable challenges in global public health, standing as a leading cause of cancer-related deaths worldwide. Besides, the relationship between chronic respiratory conditions, such as pneumonia, and the development of lung cancer highlights the necessity for advanced diagnostic tools. Prolonged inflammation caused by severe pneumonia can lead to cellular damage and

genetic mutations, creating a microenvironment conducive to carcinogenesis. This progression reveals the importance of early and precise detection of pulmonary abnormalities, as timely intervention could disrupt the pathological cascade leading to malignancy [1]. Accordingly, early diagnosis of lung cancer is essential for facilitating timely medical intervention, which can significantly ameliorates survival rates.

Traditional diagnostic methods rely heavily on the manual analysis of medical imaging. Therefore, these methods are labor-intensive and prone to subjectivity and error, especially

The associate editor coordinating the review of this manuscript and approving it for publication was Cheolsoo Park<sup>ID</sup>.

given the complex and heterogeneous manifestations of lung cancer [2]. Recently, deep learning (DL) models have achieved remarkable success in medical image analysis and pattern recognition [3], [4], driving a paradigm shift in the field, particularly for lung cancer detection [5], [6]. Such as, DL algorithms hold transformative potential of identifying subtle nodules and lesions in medical images which redefines traditional screening methodologies. By exploiting the power of deep neural networks, these advanced models can augment the diagnostic capabilities of radiologists and provide reliable second opinions or preliminary assessments that enhance efficiency in clinical workflows [7]. Convolutional neural networks (CNN), in particular, has emerged as a vital solution to autonomously identify intricate patterns and subtle anomalies within medical images, eliminating the need for domain-specific expertise. This technological evolution has not only been hailed as a groundbreaking advancement but has also fundamentally altered the landscape of diagnostic practices. Combined with access to large-scale datasets, DL has the capacity to transform the efficacy of lung cancer diagnosis and offer a preventive approach to managing respiratory ailments [8].

A growing body of research has explored the application of DL models to improve lung cancer detection using imaging modalities, including X-rays [9], [10], [11] and computed tomography (CT) scans [12], [13], [14] as well as histopathology images [15], [16]. These studies have demonstrated the ability of machine learning to extract features and complex patterns from medical images, often surpassing traditional methods by enabling the identification of malignant lesions that might be overlooked during manual examination [17]. Additionally, DL models have shown exceptional performance in detecting small pulmonary nodules, which are critical for early-stage lung cancer diagnosis [18]. By integrating these technologies into clinical practice, healthcare systems can significantly reduce diagnostic delays and improve patient outcomes.

## A. LITERATURE REVIEW

Several recent studies have contributed to advancements in lung cancer detection. Rajpurkar et al. [17] pioneered the application of DL for pneumonia detection in chest X-rays, setting a benchmark for future research. For instance, Bhandary et al. [11] developed specialized model for distinguishing between pneumonia and normal X-ray images by integrating AlexNet architecture with handcrafted features. Similarly, Alshmrani et al. [9] employed VGG19 architecture along with CNN for feature extraction from X-ray images. They attained an accuracy of 93.75%. In parallel, Bharati et al. [10] proposed a hybrid DL framework that combined VGG with a CNN, achieving a low diagnostic accuracy of 73%. Ayan et al. [19] explored the difficulties of diagnosing pneumonia by employing an ensemble of seven pretrained CNNs for distinguishing between normal, viral and bacterial pneumonia in X-ray images. Goyal and

Singh [20] presented a comprehensive framework combining image enhancement, region of interest extraction and advanced classifiers like artificial neural network (ANN), support vector machines (SVM) and long short-term memory (LSTM). This approach demonstrated superior performance in differentiating pneumonia and COVID-19, showcasing the synergy between image processing and machine learning.

On the other hand, CT scans have also been widely utilized in existing approaches for lung cancer diagnosis. Lakshmanprabu et al. [12] introduced an automated system for analyzing lung CT images, exploiting deep neural network paired with linear discriminant analysis to classify nodules as malignant or benign. They obtained an overall accuracy of 94.56%. While, Shah et al. [13] proposed an ensemble approach that combined multiple CNNs for lung nodule detection, utilizing the LUNA 16 dataset. Their model achieved an accuracy of 95% in detecting cancerous images. Rehman et al. [18] designed a CNN-based model for tumor and nodule identification, incorporating preprocessing with filtering for image enhancement and postprocessing with morphological operators for precise segmentation. Utilizing an active contour algorithm, their model achieved a dice similarity of 98.18%. To overcome the scarcity of annotated CT scans, Xie et al. [21] proposed a three-level optimization method that employs source domain data to enhance model performance. While in Wang et al. [22], a DL algorithm was developed to classify pneumonia types using CT images, achieving AUC values of 93.4%, 81.6% and 71.5% for bacterial, viral and fungal pneumonia, respectively. Notably, their model outperformed radiologists in distinguishing bacterial and viral cases.

Rajasekar et al. [14] evaluated several DL architectures, including ResNet50, VGG16, VGG19 and InceptionV3, analyzing both CT scans and histopathological images. Interestingly, their findings revealed that histopathological tissue analysis yielded higher detection accuracy compared to CT-based methods. Onward, Masud et al. [15] developed a classification approach for five classes of lung and colon cancer using histopathological images, achieving an accuracy of 96.33%. They extracted four distinct feature sets from two image processing algorithms and then integrated them to form a robust feature set optimized for classification. Kumar et al. [23] compared handcrafted feature extraction with DL-based methods, employing classifiers like gradient boosting, multilayer perceptron, SVM and random forest (RF). Deep CNN models outperformed hand-engineered approaches by approximately 5%, with DenseNet-121 paired with RF achieving the highest accuracy of 98.6%. In another study, Togacar [16] utilized DarkNet19 for training on lung and colon cancer dataset. They exploited meta-heuristic algorithms and SVM for feature selection and classification, respectively. They achieved an overall accuracy of 99.69% with a 70-30 training-testing split. Talukder et al. [24] introduced an ensemble deep feature extraction model, achieving 99.05% accuracy for lung images, 100% for colon images and an overall 99.30% for five class classification.

In contrast, Mehmood et al. [25] introduced a computationally efficient model, yielding accuracy of 98.4%. Moreover, attention mechanisms have significantly advanced medical imaging [26]. Cha et al. [27] combined transfer learning with attention-based feature selection using architectures such as ResNet152, DenseNet121 and ResNet18. Similarly, Wang et al. [28] enhanced DenseNet by incorporating a squeeze-and-excitation block to improve feature representation through focusing on lesion-specific regions. These innovations demonstrate the potential of integrating DL with attention mechanisms in healthcare diagnostics across various imaging modalities.

## B. LITERATURE REVIEW GAPS

Despite significant progress in lung cancer detection models, they suffer from several limitations that affect their clinical applicability. One of the primary challenges is the computational complexity of the existing DL methods. As previous studies employed multi-layer CNNs, hybrid frameworks and ensemble learning [10], [13], [19], [24]. However, such frameworks require substantial resources, making them impractical for deployment in resource-limited settings. Besides, techniques such as transformers and hyperspectral imaging further exacerbate this issue due to high-dimensional feature extraction and processing demands [14], [21], [27]. Accordingly, real-time implementation remains an issue, as these models often demand high computational complexity. Many studies prioritize achieving high accuracy in controlled settings, but fail to optimize models for practical healthcare settings. Without addressing these limitations, the transition of DL models from research to real-world lung cancer diagnostics is a critical challenge. Additionally, many existing models relied on extensive preprocessing steps, including feature selection and dimensionality reduction [12], [16], [20]. These additional steps introduce manual or semi-automated processes, making clinical integration more cumbersome. Moreover, although traditional feature extraction techniques followed by machine learning classification are effective on curated datasets [11], [15], [16], they may struggle to capture sophisticated lung cancer patterns in heterogeneous medical images. This reliance on predefined feature engineering limits the adaptability to unseen cases and affect the diagnostic reliability.

Generalizability is a major concern, as the majority of previous studies focused on specific datasets. As they often utilize only one dataset without external validation on other datasets [11], [12], [14]. In addition, most of these studies are based on single-modality data, such as X-rays, CT scans, and histopathological images. Additionally, many studies are constrained by the use of small datasets, which hinders the robustness and scalability of their models. Moreover, these studies typically address a single disease and therefore fail to develop comprehensive models capable of detecting multiple conditions simultaneously. As a result, they did not explore the performance of their developed methods beyond their controlled experimental settings.

## C. MOTIVATION AND OUR METHOD

Lung cancer is one of the most lethal malignancies in the world, which requires a rapid and precise diagnosis for effective treatment. Conventional screening methods depend on manual interpretation of medical images. Nevertheless, this task is labor-intensive, time-consuming and open the door for subjective evaluation. Therefore, DL which is capable of automatically identifying important biomarkers in radiological and histopathological images, has emerged as a pivotal tool to fill these gaps. However, existing DL models for lung cancer detection face significant limitations, including reliance on single imaging modalities, insufficient datasets, suboptimal feature extraction techniques and limited generalizability. Moreover, existing systems concentrate on merging features from multiple CNNs without adequate feature reduction, resulting in overly complex models that require vast amounts of data to learn effectively. The scarcity of large datasets often hampers performance, as smaller datasets struggle to capture critical patterns. Furthermore, many existing CNN models are computationally heavy, increasing the risk of overfitting. Also, despite that computationally efficient models have been introduced, they still do not have the ability to capture multi-scale informative features, which are essential for precise diagnosis. In recent developments, attention mechanisms have been embedded into CNNs to boost feature extraction capabilities. They enabled the developed models to prioritize crucial information and filter out unnecessary details. This integration has led to notable advancements in the performance of conventional DL models.

Accordingly, we introduce a novel DL framework that incorporates diverse imaging modalities, employing an attention-based modules to enhance diagnostic accuracy. Our framework is the first to apply a unified approach across mega-sized dataset comprising three distinct modalities (i.e. X-ray, CT and histopathological images). Additionally, the proposed system addresses multiple diseases, including pneumonia, lung and colon cancer. The developed architecture combines multi-scale convolutional fusion with an efficient channel attention (ECA) mechanism. In particular, the model employs Inception modules embedded with residual blocks for spatial feature extraction, followed by a hierarchical feature extraction (HFE) block to process information at multiple receptive fields which enhances feature diversity. The ECA module dynamically weights feature importance to help the model to focus on highly discriminative regions within the input images. Therefore, unlike traditional CNN-based models, our approach excels at extracting the most relevant features at multiple spatial scales. This combination guarantees that the model is capable of extracting intricate details and contextual information. To enhance model transparency and interpretability, we employed a combination of Gradient-weighted class activation mapping (Grad-CAM) and attention map visualizations to highlight the spatial regions and feature channels most influential in the model predictions. Besides, we systematically evaluated

the contribution of each component in the proposed model to validate their effectiveness. Furthermore, comparative experiments against conventional architectures in lung cancer detection have been conducted. Finally, we implemented an ablation study to gauge the generalizability of our framework. We tested the proposed model on four distinct datasets representing three different modalities, which differ in terms of features and classification challenges. These experiments are vital to test the model ability to extract informative features and maintain high performance across diverse medical images. Importantly, this versatility highlights the potential of our framework to serve as a clinically acceptable tool for a wide range of applications. By integrating multiple imaging modalities and advanced attention mechanisms, our framework has the potential to improve diagnostic accuracy. Fig. 1 provides a graphical representation of the proposed system.

#### D. NOVELTIES AND CONTRIBUTIONS

The key innovations and contributions of this study are as follows:

- A novel DL framework is introduced for lung cancer detection, integrating multiple imaging modalities to enhance diagnostic generalizability.
- The model exploits a hybrid Inception-ResNet architecture, combining the feature extraction capabilities of Inception modules with the gradient flow optimization of residual connections to enhance feature representation.
- The proposed model employs HFE modules to capture multi-scale spatial features and facilitate the detection of complex patterns across diverse images.
- ECA blocks are embedded to dynamically adjust feature importance and ensure that the model focuses on the highly discriminative regions.
- The proposed system eliminates the need for manual segmentation and extensive preprocessing, which streamlines the diagnostic process.
- The model combines Grad-CAM and attention map visualizations to provide both class-specific and feature-level insights, enhancing transparency and enabling clinicians to understand not only what the model predicts but also why and where it focuses during decision-making.
- The model achieved accuracy rates of 95.35%, 99.68%, 99.73% and 99.26% using X-ray, CT, histopathological and mega datasets, respectively.
- Comparative experiments against established CNN models demonstrate superior performance in lung cancer detection.
- Unlike many existing models that focus on a single disease, this framework has the potential to be extended for multi-disease diagnosis.

The structure of this paper is as follows: Section II describes the datasets, proposed model architecture and

TABLE 1. Details of the used dataset.

Class	Train	Validation	Test	Total
Normal	1341	8	234	1583
Pneumonia	3875	8	390	4273
Total	5216	16	624	5856

experimental setup. Section III presents the model performance, including comparisons with existing methods, and validation on additional datasets. Finally, Section IV summarizes the findings and provides recommendations for future research.

## II. METHODOLOGY

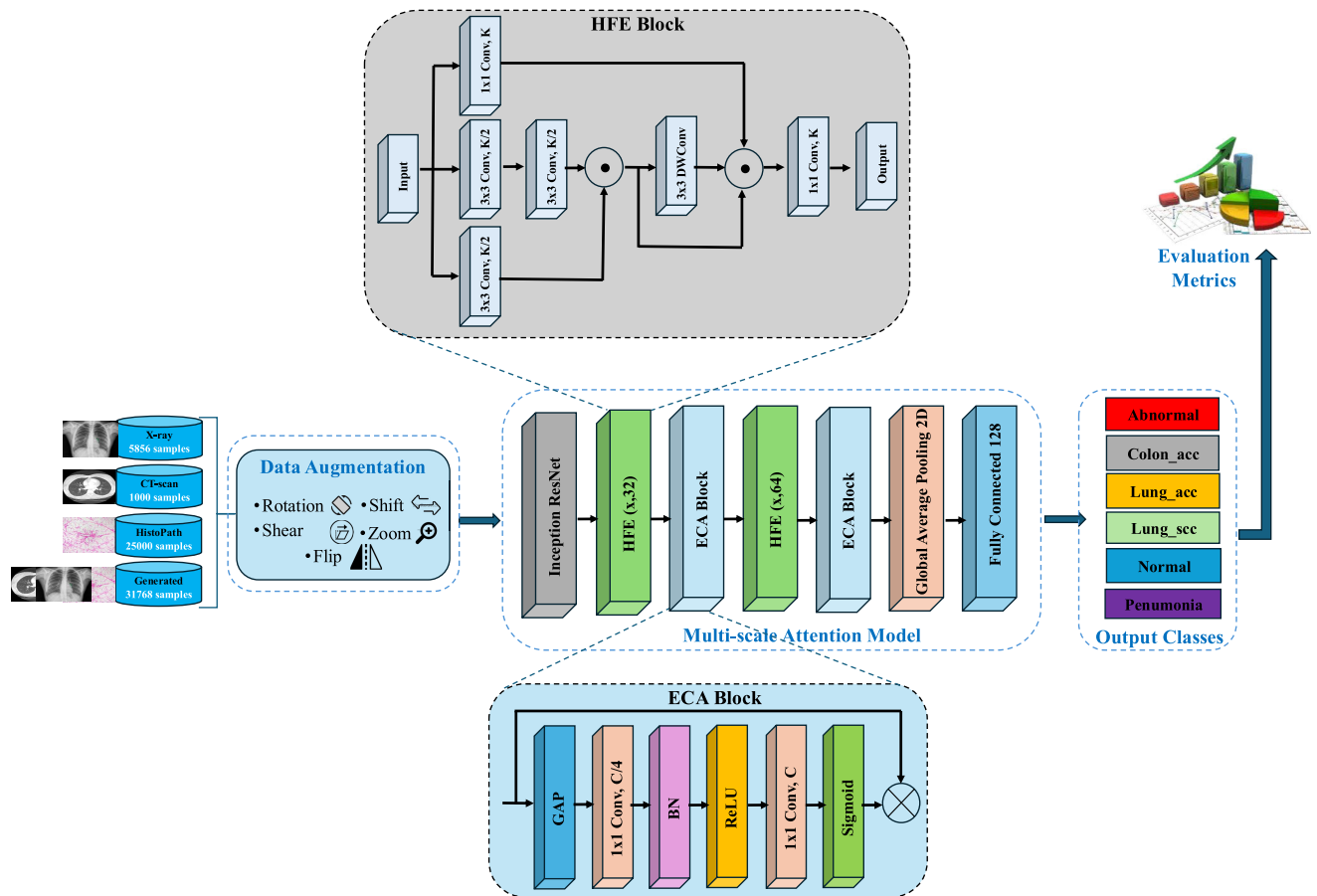
### A. DATASET

The dataset used in this study consists of chest X-ray images from pediatric patients aged one to five years, collected from the Guangzhou women and children medical center. It contains 5,856 images, categorized into pneumonia and normal cases. The dataset was sourced from Kaggle repository [29], where images were structured into train, validation and test directories. The X-ray images were provided in JPEG format, with stringent quality control measures applied to exclude low-quality scans. The employed dataset was reviewed by two physicians with an additional expert evaluation. Table 1 presents a comprehensive overview of the dataset, including the distribution and quantity of images across different categories. For model training, 5,232 images were used, comprising 3,883 pneumonia cases and 1,349 normal cases. The trained model was then evaluated on 624 unseen patient cases, including 234 normal and 390 pneumonia images. Challenges such as class imbalance and varying image quality caused by differences in scanning devices, operator techniques and patient positioning necessitated augmentation techniques. Therefore, to enhance model generalization and prevent overfitting, methods such as flipping, shifting, shearing, zooming and rotation were employed. These techniques also help to expand the dataset to better reflect real-world variability in clinical X-ray scans. Representative chest X-ray images from this dataset are shown in Fig. 2, illustrating the variability in image quality and patient conditions.

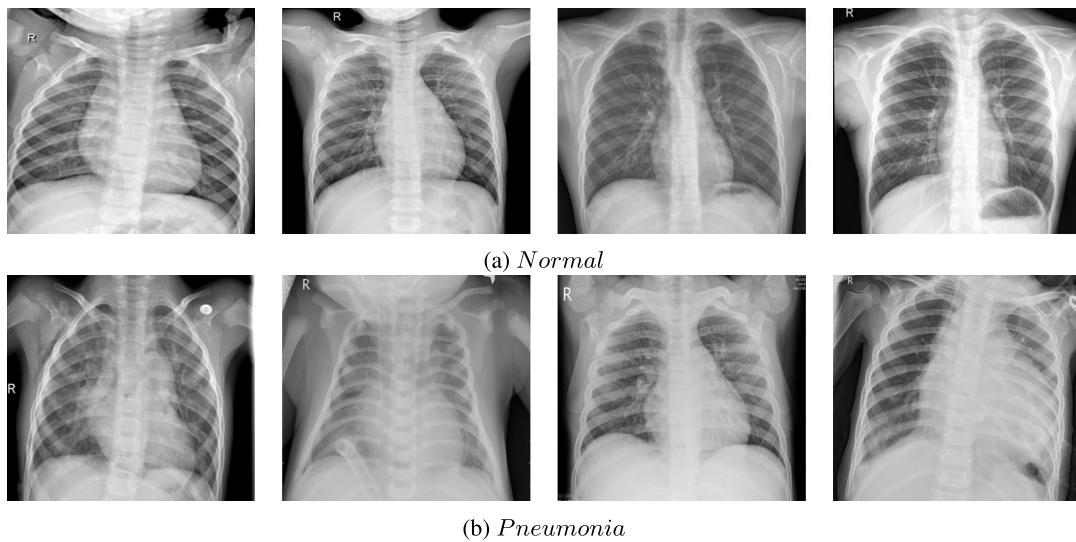
### B. PROPOSED MODEL

A multi-scale attention-based DL framework that integrates multiple imaging modalities is developed herein for lung cancer diagnosis. The methodology is divided into several key components, each designed to mitigate specific challenges in medical image analysis. The architecture is inspired and built upon the Inception and ResNet backbones. These two powerful modules blend the strengths of Inception and residual connections to extract hierarchical feature representations efficiently. To further ameliorate the model ability to recognize complex patterns in lung cancer images, we use two sequential key components namely, HFE and ECA. The multi-scale convolutional fusion in





**FIGURE 1.** Block diagram of the proposed system.



**FIGURE 2.** Sample images of normal and pneumonia type from X-ray dataset.

HFE aids the model to learn features at different scales, improving its sensitivity to both global and local structural variations in cancerous tissues. Additionally, the HFE block

strengthens feature extraction by progressively integrating spatial and channel-wise contextual information to maintain a comprehensive understanding of the lung abnormal regions.

The ECA mechanism refines feature maps by adaptively emphasizing the most informative channels, thereby improving the representation of critical patterns. Table 2 presents the detailed structure of the proposed model. Below, we provide an in-depth explanation of each architectural component.

### 1) INCEPTION-ResNet MODULE

The Inception-ResNet module is a cornerstone of our framework, designed to extract spatial features from medical images. This module takes advantage of the strengths of Inception modules and ResNet blocks by enabling the extraction of both local and global patterns while mitigating the vanishing gradient problem through residual connections. The developed module employs several convolutional pathways with varying kernel sizes to capture diverse patterns and identify subtle abnormalities in medical images. Let  $X \in \mathbb{R}^{H \times W \times C}$  denote the input image, where  $H$ ,  $W$ , and  $C$  represent height, width and the number of channels, respectively. The Inception-ResNet module processes the input data through four parallel operations. First, a  $1 \times 1$  convolution extracts local features while reducing dimensionality, expressed as  $F_1(X) = W_1 * X + b_1$ , where  $W_1$  and  $b_1$  represent the weights and biases of the  $1 \times 1$  convolutional layer. Second, a  $3 \times 3$  convolution captures medium-range spatial dependencies, defined as  $F_2(X) = W_2 * X + b_2$ , where  $W_2$  and  $b_2$  are the weights and biases of the  $3 \times 3$  convolutional layer. Third, a  $5 \times 5$  convolution extracts broader contextual information, formulated as  $F_3(X) = W_3 * X + b_3$ . Finally, maximum pooling preserves spatial invariance by downsampling the input, denoted as  $F_4(X) = \text{MaxPool}(X)$ . The outputs of these four operations are concatenated along the channel dimension and combined with a residual connection to produce the final output of the Inception-ResNet module:

$$F_{(X)} = \text{Concat}(F_1(X), F_2(X), F_3(X), F_4(X)) + X \quad (1)$$

The use of  $1 \times 1$  convolutions for dimensionality reduction within the Inception-ResNet modules reduces computational complexity. While, the addition of skip connections in each Inception-ResNet block provides efficient training of deeper networks and improves feature propagation. Moreover, residual connections ameliorates the model ability to learn hierarchical features.

### 2) RESIDUAL CONNECTIONS IN INCEPTION-ResNet

Residual connections are a critical component of the Inception-ResNet architecture to address the vanishing gradient problem. These connections are embedded into each Inception module to form Inception-ResNet blocks. Each block refines the feature representations by combining the outputs of multi-scale convolutional pathways with the original input through a skip connection. Let  $X$  denote the input to an Inception-ResNet block. The output is computed as  $F_{\text{Inception-ResNet}}(X) = \mathcal{F}(X, \{W_i\}) + X$ , where  $\mathcal{F}(X, \{W_i\})$  represents the transformation applied by the parallel convolutional pathways, and  $\{W_i\}$  denotes the set of learnable

parameters. The skip connection ensures that the gradient can flow directly through the network. The Inception-ResNet blocks are stacked to form a deep network to enable the model to learn hierarchical features with increasing levels of abstraction. This architecture is particularly effective for medical image analysis, where fine-grained details are important for accurate diagnosis. By integrating residual connections into the Inception modules, the model can achieve superior performance while maintaining computational efficiency.

### 3) HIERARCHICAL FEATURE EXTRACTION BLOCK

The HFE Block is a sophisticated module designed to enhance feature diversity by processing information at multiple receptive fields to ensure the retention of discriminative information. This block employs parallel convolutional paths and hierarchical concatenation to capture local, medium-range and broader contextual information. By combining features from multiple scales, the HFE block advances the model ability to discriminate between complex patterns in the X-ray images. Let  $X \in \mathbb{R}^{H \times W \times C}$  denote the input feature map, where  $H$ ,  $W$ , and  $C$  represent the height, width, and number of channels, respectively. The HFE block processes the input through multiple convolutional layers and combines their outputs hierarchically to produce a rich feature representation.

The proposed HFE block comprises three distinct convolutional pathways. The first one employs a  $1 \times 1$  convolution, which is primarily responsible for extracting local features while simultaneously reducing dimensionality. While, pathway 2 consists of a hierarchical structure with two sequential  $3 \times 3$  convolutions, allowing the model to capture medium-range spatial dependencies within the feature map. Pathway 3 applies a single  $3 \times 3$  convolution, designed to extract broader contextual information from the input feature map. After feature extraction through these three pathways, the outputs are concatenated and fused using additional convolutional layers to create a comprehensive feature maps. The fusion process is mathematically defined as:

$$F_{\text{HFE}}(X) = W_f * \text{Concat}(F_1(X), F_2(X), F_3(X)) + b_f \quad (2)$$

where  $W_f$  and  $b_f$  correspond to the weights and biases of the final fusion layer. This hierarchical approach helps the model to extract multi-scale spatial information to differentiate between fine-grained and high-level features. Additionally, it utilizes efficient convolutional operations, minimizing computational overhead while maintaining high feature diversity. By refining learned representations, the HFE block significantly improves the model ability to distinguish subtle patterns in medical images, making it a powerful enhancement for DL-based medical diagnostics. Algorithm 1 illustrates the HFE implementation in our study.

### 4) EFFICIENT CHANNEL ATTENTION MECHANISM

ECA is a lightweight attention mechanism designed to dynamically recalibrate the importance of feature channels.

**TABLE 2.** Details of the proposed model components.

Layers	Description	Output Shape	Params.
Input layer	-	$224 \times 224 \times 3$	0
Inception-ResNet	20 stacked Inception blocks, Residual learning via identity shortcuts	$5 \times 5 \times 1536$	54,275,232
HFE block	$[1 \times 1Conv], [3 \times 3Conv] \times 2$ , Swish activation	$5 \times 5 \times 32$	733,216
ECA block	Global average pooling, $[1DConv(k = 32)]$ , Sigmoid, Channel-wise multiplication	$5 \times 5 \times 32$	32
HFE block	$[1 \times 1Conv], [3 \times 3Conv] \times 2$ , Swish activation	$5 \times 5 \times 64$	110,144
ECA block	Global average pooling, $[1DConv(k = 32)]$ , Sigmoid, Channel-wise multiplication	$5 \times 5 \times 64$	32
GAP	Global average pooling	$1 \times 1 \times 64$	0
FC layer	128 units, ReLU activation	$1 \times 1 \times 128$	8320
Classifier	2 units, Softmax activation	$1 \times 1 \times 2$	258

**Algorithm 1** Hierarchical Feature Extraction Block

- 1: **Input:** A feature map  $X \in \mathbb{R}^{H \times W \times C}$ .
- 2: **Pathway 1:** Apply a  $1 \times 1$  convolution.
- 3: **Pathway 2:** Apply two sequential  $3 \times 3$  convolutions.
- 4: **Pathway 3:** Apply a single  $3 \times 3$  convolution.
- 5: **Concatenation:** Concatenate the outputs of Pathways 2 and 3.
- 6: **Fusion:** Apply a  $3 \times 3$  convolution to the concatenated output.
- 7: **Final concatenation:** Concatenate the fused output with the output of Pathway 1.
- 8: **Output:** Apply a  $1 \times 1$  convolution to the final concatenated output.

The implemented ECA mechanism increases the model ability to emphasize discriminative regions within the X-ray images. This mechanism is particularly advantageous for medical image analysis, where fine-grained feature discrimination is paramount. Let  $X \in \mathbb{R}^{H \times W \times C}$  denote the input feature map, where  $H$ ,  $W$ , and  $C$  represent the height, width and number of channels, respectively. The ECA mechanism operates in three key steps. The first step aggregates spatial information from the input feature map to produce a channel-wise descriptor. For each channel  $c$ , the global average pooling (GAP) operation computes the average value across the spatial dimensions as follows:

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X(i, j, c) \quad (3)$$

where  $z_c$  represents the channel-wise descriptor for the  $c$ -th channel. This step reduces the feature map  $X$  to a vector  $z \in \mathbb{R}^C$  to extract the global spatial context. To extract cross-channel interactions without increasing computational complexity, a 1D convolution is applied to the channel-wise descriptor  $z$ . The convolution uses a kernel size  $k$  to aggregate information from neighboring channels as in the following equation:

$$s = \sigma(W_{1D} * z) \quad (4)$$

where  $W_{1D}$  denotes the weights of the 1D convolutional layer,  $*$  represents the convolution operation, and  $\sigma$  is the sigmoid activation function. The output  $s \in \mathbb{R}^C$  represents the attention weights for each channel. The final step calculates the input feature map  $X$  by multiplying each channel with the corresponding attention weight as in the following equation:

$$\hat{X}(i, j, c) = s_c \cdot X(i, j, c) \quad (5)$$

where  $\hat{X}$  is the weighted feature map. This operation ameliorates the contribution of important channels while suppressing less relevant ones. The ECA mechanism can be summarized in algorithm 2. The ECA mechanism

**Algorithm 2** Efficient Channel Attention Mechanism

- 1: **Input:** Feature map  $X \in \mathbb{R}^{H \times W \times C}$
- 2: Compute the channel-wise descriptor  $z \in \mathbb{R}^C$  by averaging spatial dimensions:

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X(i, j, c) \quad (6)$$

- 3: Apply a 1D convolution with kernel size  $k$  to capture cross-channel interactions, followed by a sigmoid activation to generate attention weights  $s \in \mathbb{R}^C$ :

$$s = \sigma(W_{1D} \cdot z) \quad (7)$$

- 4: Multiply the input feature map  $X$  by the attention weights  $s$  to produce the recalibrated feature map  $\hat{X}$ :

$$\hat{X}(i, j, c) = s_c \cdot X(i, j, c) \quad (8)$$

- 5: **Output:** The recalibrated feature map  $\hat{X}$ , which emphasizes discriminative channels.

improves feature representation while introducing minimal computational overhead compared to traditional attention mechanisms, as it avoids the use of fully connected layers. Instead, it employs 1D convolution, which efficiently models cross-channel dependencies without significantly increasing the parameter count. The ECA mechanism ensures that the model focuses on diagnostically significant patterns for accurate classification.

### 5) MODEL TRAINING AND EVALUATION

The training process is designed to optimize the model performance and ensure scalability across diverse datasets. Images are resized to  $224 \times 224$  pixels. To enhance the model generalization, a comprehensive set of data augmentation techniques are employed. First, random rotations within a range of  $\pm 15^\circ$  are applied to simulate variations in orientation. Additionally, random horizontal and vertical shifts within 10% of the image dimensions helped to account for positional variability. To introduce geometric distortions, random shearing with an intensity of 0.1 is incorporated. The augmentation pipeline also included random zooming within a range of 80% to 120% to simulate variations in scale. Furthermore, random horizontal flipping is performed to account for symmetry in medical images. Finally, any empty pixels resulting from these transformations are filled using the nearest neighbor method to preserve image integrity. The developed model is trained using the categorical cross-entropy loss. This loss function is well-suited for multi-class classification tasks. The training hyperparameters, including batch size, number of epochs, and learning rate configuration as well as the architectural design choices such as kernel sizes, number of layers, and attention configurations, were selected empirically through iterative experimentation and validated based on performance across the employed dataset. Adam optimizer with batch size of 20 and initial learning rate of 0.0001 is utilized. To further improve training stability and convergence, we employed adaptive learning rate scheduling. The learning rate was reduced by a factor of 0.5 if the validation loss does not improve for 2 consecutive epochs, with a minimum learning rate of  $1 \times 10^{-6}$ . The model was trained for 20 epochs. The proposed approach was implemented using the Keras framework, with TensorFlow serving as the backend. Experiments were carried out on a system featuring an NVIDIA Tesla P100 GPU and 32 GB of RAM. Performance assessment relied on critical metrics such as accuracy, specificity, sensitivity, precision, area under the curve (AUC) and F1-Score.

## III. RESULTS AND DISCUSSION

This section presents a thorough assessment of the introduced model. It starts with an extensive examination of the performance of the proposed architecture. Key contributions of each component in the developed model are analyzed. Additionally, the framework effectiveness is compared to leading CNN-based approaches, alongside prior studies on lung cancer detection. Finally, to further gauge the adaptability of the proposed method, we conducted additional evaluations across three distinct datasets. This comprehensive assessment reveal the proposed system generalizability across diverse data scenarios.

### A. THE OBTAINED RESULTS USING X-RAY DATASET

Table 3 illustrates the performance of the proposed model on the X-ray dataset. The model exhibits a strong ability

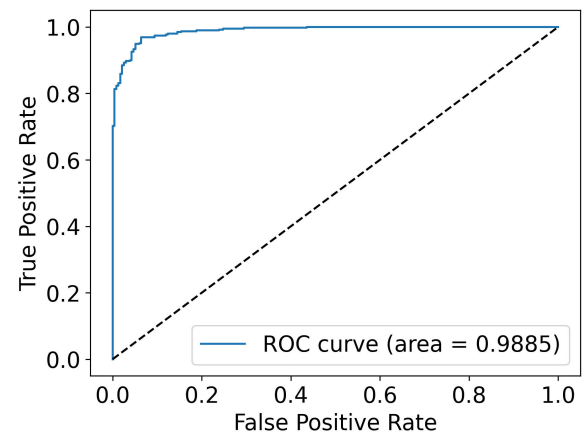


FIGURE 3. ROC curve of the proposed model.

to distinguish between normal and pneumonia cases with an overall accuracy of 95.35%. Notably, the balanced precision, F1-score, sensitivity and specificity across both classes indicate that the model does not favor one class over the other. For the normal class, the model attained a precision of 93.99%, ensuring that healthy patients are not incorrectly diagnosed. Sensitivity and specificity values reinforce this observation, showing that the model is adept at detecting true normal cases while successfully rejecting pneumonia cases. Similarly, for the pneumonia class, the model demonstrated a precision of 96.16% and F1-score of 96.29%, ensuring that pneumonia cases are accurately detected. Meanwhile, the specificity of 93.59% confirms that the model is effective at excluding normal cases from being misdiagnosed as pneumonia. The receiver operating characteristic (ROC) curve presented in Fig. 3 illustrates the excellent performance of the proposed model in distinguishing between pneumonia and normal cases. The model achieves an AUC of 0.9885. The high AUC value reveals that the model maintains a strong balance between sensitivity and specificity. Unlike traditional models, which often struggle with imbalanced sensitivity and specificity, our approach maintains consistent classification performance across both categories, making it a valuable tool for efficient medical diagnosis.

To enhance the transparency of our model, we incorporated visual explanation techniques using both attention maps and Grad-CAM visualizations. These methods illustrate which regions the model considers most important when making diagnostic decisions. For the attention-based visualizations, we extracted feature maps from the final ECA mechanism. The output of this block was visualized and averaged spatially to generate heatmaps that reflect where the model learned to focus during training. These attention maps are not class-specific but reflect generalized discriminative areas learned during feature extraction. The Grad-CAM heatmaps were generated post hoc using gradient-based localization. Specifically, we computed the Grad-CAM using the last Convolutional layer immediately preceding the final ECA



**TABLE 3.** Testing results of the proposed model using X-ray dataset.

Class	Accuracy (%)	Precision (%)	F1-Score (%)	Sensitivity (%)	Specificity (%)
Normal	95.35	93.99	93.79	93.59	96.41
Pneumonia	95.35	96.16	96.29	96.41	93.59

block. The gradients of the predicted class score was calculated with respect to the corresponding convolutional feature maps. These heat maps provide class-specific localization, identifying spatial regions that contributed the most to the final predicted class. Figure 4 shows representative examples of correctly classified cases for the normal and pneumonia classes. In normal cases, both the attention and the Grad-CAM heatmaps consistently emphasize the central lung fields, particularly the midzones of the thorax. These are the anatomical regions that radiologists routinely examine for symmetry, clarity, and the absence of infiltrates. The attention of the model to these clinically relevant areas suggests that it has learned to focus on well-aerated, structurally intact lung regions. In contrast, in the pneumonia cases, both the attention and Grad-CAM heat maps clearly shift their focus toward dense, asymmetrical regions of the lung fields, most notably in the lower and middle lobes, which are common sites for pneumonia consolidation. The high-activity zones on these maps correspond to visual features such as increased opacities and parenchymal distortion, which are hallmark signs of pulmonary infection. The close match between these highlighted regions and the known radiological manifestations of pneumonia indicates that the model is not just learning superficial features, but is internally attending to structurally and diagnostically significant cues. Importantly, the ECA attention maps show the model ability to emphasize clinically relevant features early on, with their alignment to Grad-CAM outputs. This convergence of visual explanation methods confirms the model intelligent localization ability despite being trained without any bounding box or segmentation supervision.

### B. PERFORMANCE OF DIFFERENT COMBINATIONS OF THE PROPOSED MODEL

Figure 5 displays the progressive improvement in classification performance as different architectural components are incorporated into the implemented model. By analyzing these matrices, we observe how each additional component refines the model ability to extract meaningful features and reduce misclassification rates. Fig. 5(a) depicts the performance of the Inception model alone. Despite that the model can effectively capture spatial hierarchies in images, feature extraction capabilities remain limited which led to 40 normal cases being misclassified as pneumonia. The model successfully detects 385 pneumonia cases, misclassifying only 5 as normal. However, the higher false positive rate signifies that the model struggles to differentiate variations between the two classes. ResNet blocks are incorporated, forming

the Inception-ResNet model in Fig. 5(b). The addition of residual connections significantly enhanced feature propagation. As a result, the number of correctly classified normal cases increases from 194 to 214. However, this adjustment introduces a slight trade-off, as 15 pneumonia cases are now misclassified as normal. Despite this, the overall performance improves, demonstrating that residual connections produce more effective learning of complex features.

Building upon the Inception-ResNet framework, the integration of an ECA mechanism in Fig. 5(c) further enhanced feature discrimination by adaptively prioritizing the most relevant channels. This targeted focus allows the model to emphasize the most informative regions in the input images. The number of correctly classified pneumonia cases increases to 381, reducing the misclassification of pneumonia as normal from 15 to 9. Fig. 5(d) represents the proposed model, which merged HFE blocks alongside the previous components. This addition provides a multi-level feature learning strategy to capture high-level representations. As a result, the model achieves the highest classification accuracy, with 219 normal cases correctly identified and only 15 misclassified, marking the lowest false-positive rate among all models. The HFE block ameliorated the ability to generalize across diverse X-ray images patterns. This gradual improvements highlight the importance of integrating advanced feature extraction techniques in medical image classification. To accentuate, while the Inception model serves as a strong baseline, the addition of ResNet connections enhanced feature propagation, ECA improved attention to critical regions and HFE optimized hierarchical learning.

### C. PERFORMANCE COMPARISON WITH CNN TECHNIQUES

Table 4 and Fig. 6 show a comprehensive comparison between the proposed model and a variety of widely used CNN architectures for the classification of pneumonia. The proposed model outperformed all baseline models, achieving the highest accuracy of 95.35%, precision of 95.35%, F1-score of 95.35%, and specificity of 93.59%. Notably, while models like ResNet50 and DenseNet121 showed competitive performance, they fell short in balancing sensitivity and specificity. InceptionV3 and VGG19 demonstrated high sensitivity of 98.72% but lower specificity, indicating a tendency toward over-detection of pneumonia cases. In contrast, our model maintained a better balance, minimizing both false positives and false negatives. Additionally, recent lightweight models such as EfficientNetV2B0, MobileNetV2, and

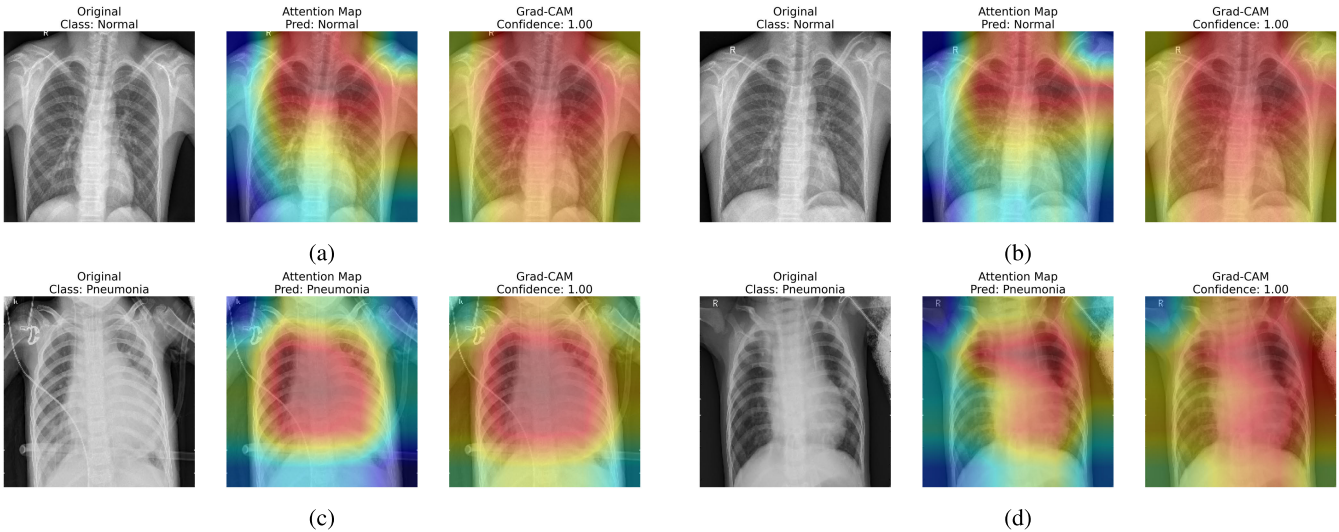


FIGURE 4. Visual explanation of the model predictions using attention maps and Grad-CAM heatmaps.

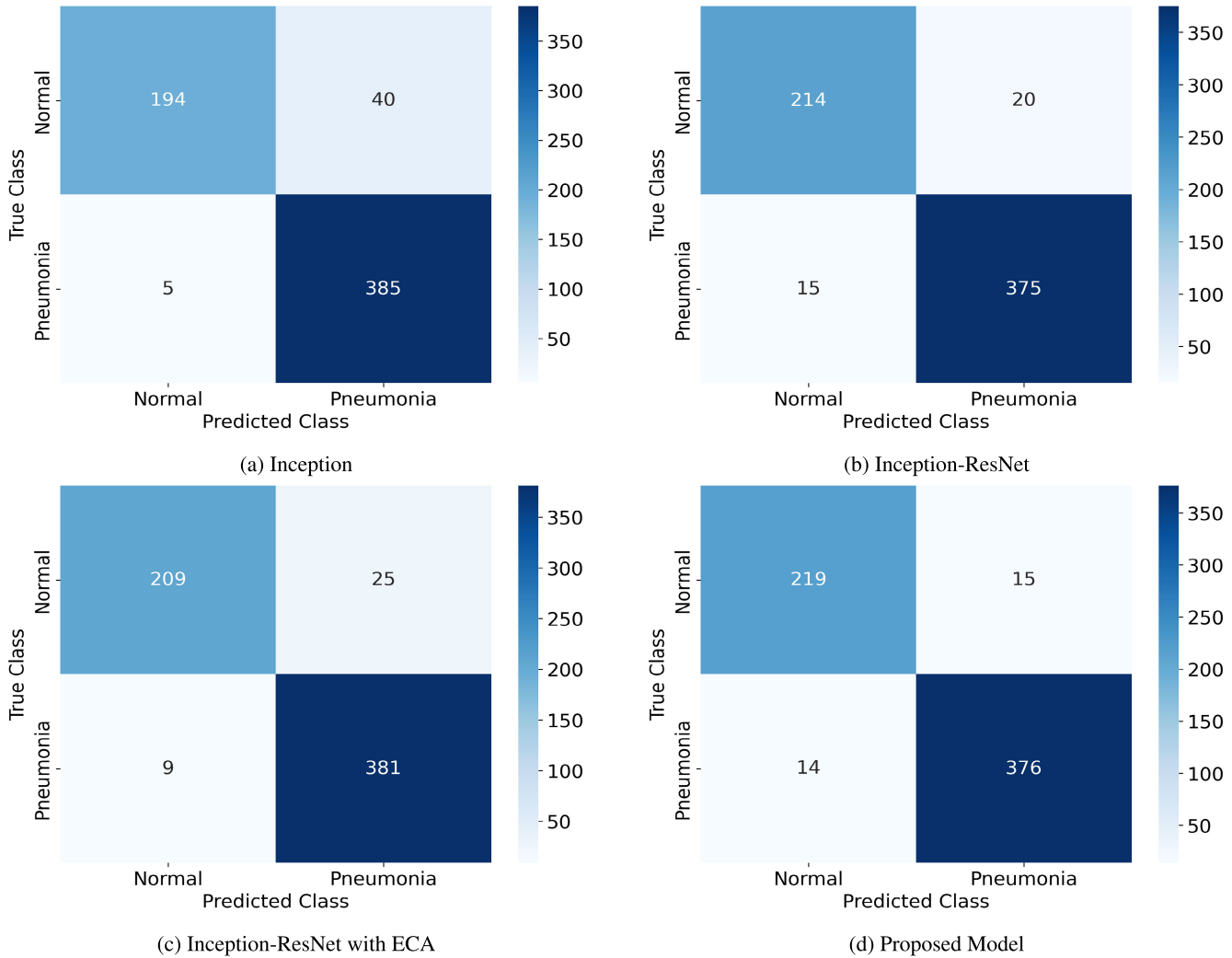
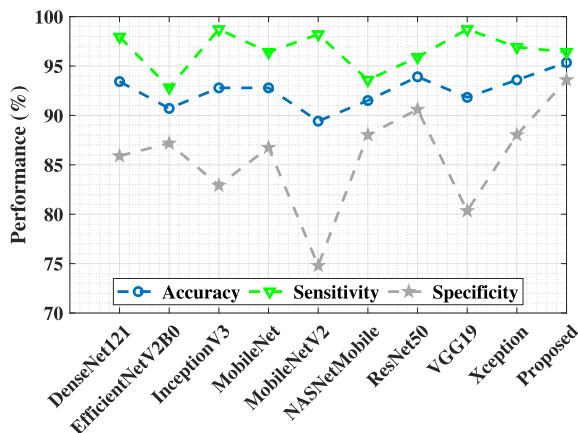


FIGURE 5. Confusion matrix obtained using different model combinations.

**TABLE 4.** Comparison of the proposed model results with longstanding CNN models.

Model	Accuracy (%)	Precision (%)	F1-Score (%)	Sensitivity (%)	Specificity (%)
DenseNet121	93.43	93.59	93.35	97.95	85.90
EfficientNetV2B0	90.71	90.69	90.70	92.82	87.18
InceptionV3	92.79	93.18	92.65	98.72	82.91
MobileNet	92.79	92.82	92.73	96.41	86.75
MobileNetV2	89.42	90.21	89.09	98.21	74.79
NASNetMobile	91.51	91.49	91.50	93.59	88.03
ResNet50	93.91	93.90	93.89	95.90	90.60
VGG19	91.83	92.36	91.64	98.72	80.34
Xception	93.59	93.63	93.54	96.92	88.03
Proposed	95.35	95.35	95.35	96.41	93.59

**FIGURE 6.** Comparison of the proposed model with CNN models.

NASNetMobile, while efficient, underperformed compared to deeper architectures and our proposed framework. The performance superiority of the proposed model justifies its architectural enhancements and supports its suitability for deployment in clinical diagnostic systems.

McNemar's  $\chi^2$  test is a non-parametric statistical method used to evaluate whether there is a significant difference in the performance of two paired classification models by specifically analyzing their discordant prediction cases [30]. Cochran's Q test serves as a generalization of McNemar's test for comparing more than two related classifiers. Once the general Q test rejects the null hypothesis, indicating that at least one of the  $k$  classifiers differs in precision, post hoc McNemar's tests can be applied to identify which specific pairs of classifiers exhibit statistically significant differences [31]. Therefore, in this study, to ascertain whether candidate classifiers exhibited statistically significant differences in predictive accuracy, Cochran's Q test was first applied to the per-instance correctness matrix, yielding a significant result ( $\chi^2 = 59.83$ ,  $df = 9$ ,  $p = 1.4 \times 10^{-9}$ ;  $p < 0.001$ ), proving that at least one model accuracy differs from the rest. Subsequent pairwise McNemar tests, adjusted for 45 comparisons using the Holm-Bonferroni correction (family-wise  $\alpha = 0.05$ ), revealed distinct performance hierarchies. Table 5 presents the results, including

the number of discordant pairs, test statistics, raw and adjusted  $p$ -values, and the significance of each comparison. The Proposed model demonstrated superiority with four statistically significant advantages over MobileNetV2 ( $p = 1.21 \times 10^{-6}$ ), EfficientNetV2B0 ( $p = 2.99 \times 10^{-5}$ ), NasNetMobile ( $p = 1.26 \times 10^{-4}$ ), and VGG19 ( $p = 0.001$ ), achieving an unmatched win, loss differential (+4). A secondary cohort, DenseNet121, InceptionV3, ResNet50, and Xception, each secured one significant victory against MobileNetV2 ( $p < 0.05$ ) without defeats, resulting in neutral differentials (+1). Intermediate models (MobileNet, EfficientNetV2B0, NasNetMobile, VGG19) exhibited no significant pairwise differences ( $p \geq 0.05$  post-adjustment), suggesting statistical equivalence. Conversely, MobileNetV2 underperformed significantly in five comparisons (vs. Proposed, DenseNet121, ResNet50, Xception, and InceptionV3;  $p < 0.05$ ), culminating in a net deficit (−5), marking it as the weakest architecture in the pool.

The comparative analysis, presented in Table 6, reveals significant insights into the performance characteristics of the implemented models. The proposed model achieved a 4.6% accuracy improvement over MobileNetV2 (89.42%) and a 1.4% gain compared to ResNet50 (93.91%), despite requiring 13.01 GFLOPs. This superior performance stems from the Inception-ResNet backbone, HFE blocks, and ECA mechanism working together to extract and enhance critical features. Where lightweight models like MobileNetV2 struggle with subtle pathological patterns due to their depthwise separable convolutions, our 20-layer Inception-ResNet module captures fine-grained textures characteristic of early-stage malignancies. The ECA mechanism proves particularly valuable in medical imaging, boosting specificity to 93.59% compared to VGG19 (80.34%) by suppressing noise in non-informative channels while amplifying features correlated with malignant findings. While the proposed model requires 55.13 million parameters ( $2.15 \times$  ResNet50) and 210.2MB storage space, this investment yields exceptional parameter efficiency. The architecture achieves an accuracy-to-parameter ratio of 1.73%/M, substantially outperforming VGG19 0.64%/M ratio. Detailed profiling reveals that 96% of parameters reside in the Inception-ResNet module, which generates hierarchical feature representations

**TABLE 5.** Results of pairwise McNemar's  $\chi^2$  tests comparing the classification performance of all model combinations.

Model 1	Model 2	Discordant (b)	Discordant (c)	McNemar $\chi^2$	p-value	p-value (adjusted)	Significant
MobileNetV2	Proposed	9	46	23.563 64	$1.208\,499 \times 10^{-6}$	$5.438\,246 \times 10^{-5}$	True
EfficientNetV2B0	Proposed	8	37	17.422 22	$2.993\,062 \times 10^{-5}$	0.001 316 947	True
DenseNet121	MobileNetV2	30	5	16.457 14	$4.976\,233 \times 10^{-5}$	0.002 139 780	True
MobileNetV2	ResNet50	9	37	15.847 83	$6.864\,535 \times 10^{-5}$	0.002 883 105	True
MobileNetV2	Xception	7	33	15.625 00	$7.722\,680 \times 10^{-5}$	0.003 166 299	True
NasNetMobile	Proposed	6	30	14.694 44	0.000 126 418	0.005 056 739	True
InceptionV3	MobileNetV2	26	5	12.903 23	0.000 328 016	0.012 792 636	True
Proposed	VGG19	32	10	10.500 00	0.001 193 745	0.045 362 327	True
MobileNet	MobileNetV2	31	10	9.756 10	0.001 787 289	0.066 129 694	False
EfficientNetV2B0	ResNet50	9	29	9.500 00	0.002 054 719	0.073 969 883	False
MobileNet	Proposed	7	23	7.500 00	0.006 169 899	0.215 946 476	False
NasNetMobile	ResNet50	6	21	7.259 26	0.007 053 638	0.239 823 688	False
MobileNetV2	VGG19	6	21	7.259 26	0.007 053 638	0.239 823 688	False
EfficientNetV2B0	Xception	13	31	6.568 18	0.010 381 796	0.332 217 465	False
InceptionV3	Proposed	10	26	6.250 00	0.012 419 331	0.384 999 250	False
DenseNet121	EfficientNetV2B0	29	12	6.243 90	0.012 462 158	0.384 999 250	False
EfficientNetV2B0	VGG19	17	24	0.878 05	0.348 736 660	1.000 000 000	False
EfficientNetV2B0	NasNetMobile	19	24	0.372 09	0.541 865 596	1.000 000 000	False
EfficientNetV2B0	MobileNet	15	28	3.348 84	0.067 252 504	1.000 000 000	False
EfficientNetV2B0	MobileNetV2	32	24	0.349 574 806	0.349 574 806	1.000 000 000	False
EfficientNetV2B0	InceptionV3	16	29	3.200 00	0.073 638 270	1.000 000 000	False
DenseNet121	Xception	8	9	0.000 00	1.000 000 000	1.000 000 000	False
InceptionV3	NasNetMobile	26	18	1.113 64	0.291 292 848	1.000 000 000	False
InceptionV3	MobileNet	13	13	0.038 46	0.844 519 267	1.000 000 000	False
InceptionV3	Xception	10	15	0.640 00	0.423 710 797	1.000 000 000	False
InceptionV3	VGG19	16	10	0.961 54	0.326 799 568	1.000 000 000	False
DenseNet121	VGG19	17	7	3.375 00	0.066 192 580	1.000 000 000	False
DenseNet121	ResNet50	9	12	0.190 48	0.662 520 584	1.000 000 000	False
DenseNet121	NasNetMobile	25	13	3.184 21	0.074 352 905	1.000 000 000	False
DenseNet121	Proposed	10	22	3.781 25	0.051 829 927	1.000 000 000	False
DenseNet121	InceptionV3	15	11	0.346 15	0.556 298 461	1.000 000 000	False
DenseNet121	MobileNet	13	9	0.409 09	0.522 431 285	1.000 000 000	False
MobileNetV2	NasNetMobile	20	33	2.716 98	0.099 285 635	1.000 000 000	False
MobileNet	Xception	7	12	0.842 11	0.358 795 358	1.000 000 000	False
MobileNet	VGG19	22	16	0.657 89	0.417 304 166	1.000 000 000	False
MobileNet	ResNet50	6	13	1.894 74	0.168 668 619	1.000 000 000	False
MobileNet	NasNetMobile	20	12	1.531 25	0.215 924 939	1.000 000 000	False
InceptionV3	ResNet50	12	19	1.161 29	0.281 198 100	1.000 000 000	False
NasNetMobile	Xception	11	24	4.114 29	0.042 522 478	1.000 000 000	False
NasNetMobile	VGG19	23	25	0.020 83	0.885 233 914	1.000 000 000	False
Proposed	ResNet50	15	6	3.047 62	0.080 855 598	1.000 000 000	False
Proposed	Xception	19	8	3.703 70	0.054 291 828	1.000 000 000	False
ResNet50	VGG19	24	11	4.114 29	0.042 522 478	1.000 000 000	False
ResNet50	Xception	9	7	0.062 50	0.802 587 349	1.000 000 000	False
VGG19	Xception	11	22	3.030 30	0.081 722 752	1.000 000 000	False

critical for differentiating between different patterns. The surprisingly modest 32-parameter ECA blocks contribute disproportionately to performance, demonstrating that strategic placement of lightweight attention mechanisms can yield 3-5% accuracy gains with negligible computational overhead. Inference time measurements reflect the computational intensity of cross-channel interactions in the ECA mechanism and multi-scale fusion operations in HFE blocks, though this remains clinically acceptable for batch processing in diagnostic workflows.

The performance gap between our model and efficient architectures like MobileNetV2 (89.42%) and EfficientNetV2B0 (90.71%) highlights fundamental limitations of general-purpose designs for medical imaging. Depth-wise separable convolutions, while computationally efficient, prove inadequate for capturing the complex textural

signatures of pulmonary lesions. Similarly, NASNetMobile demonstrates that neural architecture search optimized for natural images may not transfer optimally to medical domains. Even high-capacity models like VGG19 underperform despite their 143.67 million parameters, as their uniform  $3 \times 3$  convolution stacks lack the multi-scale processing and channel-wise adaptability of the proposed HFE-ECA modules. Interestingly, Xception achieves competitive results (93.59%) with fewer parameters (22.91M), suggesting depthwise separable convolutions can be effective when combined with residual connections, though still falling short of our model diagnostic precision. In clinical practice, 1.4-4.6% improvement over alternatives could significantly impact early cancer detection rates. Future work could optimize deployment through selective pruning of redundant Inception-ResNet branches or 8-bit quantization of ECA



**TABLE 6.** Comparison of computational efficiency metrics for the implemented models on the test set.

Model	Accuracy (%)	FLOPs (G)	Parameters (M)	Inference Time (ms)	Size (MB)
DenseNet121	93.43	5.7	8.06	61.22	32.08
EfficientNetV2B0	90.71	1.46	7.2	57.52	28.3
InceptionV3	92.79	5.7	23.85	62.43	91.95
MobileNet	92.79	1.15	4.25	58.88	16.51
MobileNetV2	89.42	0.62	3.54	53.11	14.01
NASNetMobile	91.51	1.15	5.33	65.87	22.73
ResNet50	93.91	7.76	25.64	55.31	98.37
VGG19	91.83	39.29	143.67	69.08	548.13
Xception	93.59	9.14	22.91	53.16	87.83
Proposed	95.35	13.01	55.13	79.3	210.2

layers without sacrificing accuracy. Our implementation strategically prioritizes diagnostic accuracy to ensure clinical reliability, with the model size being well within the capacity of modern medical imaging workstations.

#### D. PERFORMANCE COMPARISON WITH PREVIOUS STUDIES

Table 7 shows a comparison between the proposed model and previous studies that utilized the same chest X-ray dataset. The comparison highlights several key advantages of our model in terms of architecture efficiency, computational complexity, generalizability and classification performance. While many prior studies relied on complex ensemble models or multi-stage classification pipelines, our proposed approach streamlines the entire process by employing sequential HFE and ECA blocks within a single DL framework. This eliminates the need for additional steps such as manual feature engineering, normalization and separate classification phases as in [10], [20], and [32]. As this traditional workflow add extra computational burdens and reduce the model interpretability. Unlike these approaches, our method performs end-to-end learning, allowing the model to extract, refine and classify features in a unified framework. Besides, several studies depended solely on standard CNN architectures [33], [34] or pretrained models [32], [35] for feature extraction. Although pretrained models can effectively extract low-level and mid-level features, they did not apply any additional feature refinement techniques. Instead, while our framework is inspired by pretrained models for initial feature extraction, we introduce HFE blocks to strengthen the extracted features and make them more informative. Accordingly, our model guarantees that the extracted features are not only preserved but also optimized, leading to superior classification performance. These additional processing steps allow our model to surpass traditional feature extractors and achieve a more generalizable decision-making process.

Additionally, previous approaches combined multiple pretrained models or hybrid DL architectures, such as the study by Mabrouk et al. [36], which used a combination of Transformer, MobileNet and DenseNet169, and An et al. [5], which employed a combination of EfficientNetB0 and DenseNet121. While these ensemble methods aimed to

improve accuracy by exploiting diverse feature extraction capabilities, they inherently increase computational complexity, memory consumption and inference time, making them less suitable for online applications [38]. In contrast, our model directly learns hierarchical and spatial features through Inception-ResNet module without relying on multiple networks. Furthermore, prior studies relied on CNNs without investigating advanced attention mechanisms [6], [20], [32], [33], [34]. Our model integrates ECA with the HFE block, ensuring that not only local spatial patterns but also global contextual dependencies are effectively detected. Despite that some prior works exploited attention mechanisms [5], [28], they primarily employed basic attention mechanisms. Our model enhances attention effectiveness by integrating HFE blocks, which refine feature representations across multiple levels, making the ECA mechanism more powerful. Lack of data augmentation is another critical limitation of prior studies, which degraded their performance. Real-world clinical datasets contain noise, variations in patient positioning and differences in imaging devices. Therefore, data augmentation is crucial to elevate model generalization. Importantly, the generalizability of the majority of previous models remains questionable as their methods were tested only on a single dataset with a single imaging modality. This raises concerns about their ability to perform well on larger, more diverse datasets and different imaging modalities, such as CT scans or histopathological images. Additionally, these models were developed only for pneumonia classification. Accordingly, their applicability to other diseases, including different types of cancer or respiratory illnesses is still unverified. In contrast, our model is validated on another three datasets including three different imaging modalities.

#### E. ABLATION STUDY

To rigorously assess the generalizability of the proposed framework, we designed a multi-stage validation approach using diverse medical imaging datasets. The evaluation begins with our primary X-ray dataset, which formed the foundation for model development. We then systematically tested the framework adaptability across two additional modalities: (1) chest CT scans representing 3D volumetric imaging, and (2) histopathological images capturing

**TABLE 7.** Comparison between the state-of-the-art methods developed for lung cancer classification using the same X-ray dataset.

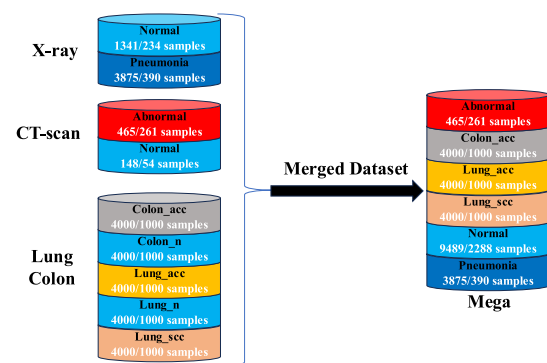
Authors	Year	Methodology	Accuracy (%)
Stephen <i>et al.</i> [33]	2019	CNN	93.73
Liang and Zheng <i>et al.</i> [34]	2020	CNN	90.50
Bharati <i>et al.</i> [10]	2020	Transformer + VGG16 + CNN	73
Salehi <i>et al.</i> [35]	2021	DenseNet121	86.80
Wang <i>et al.</i> [28]	2022	DenseNet + Attention mechanism	92.8
Mabrouk <i>et al.</i> [36]	2022	Transformer + MobileNet + DenseNet169	93.91
Goyal <i>et al.</i> [20]	2023	Feature extraction and normalization + LSTM	94.31
Bhatt <i>et al.</i> [6]	2023	CNN	84.12
Sharma <i>et al.</i> [32]	2023	VGG16 + Neural networks	92.15
Jaya and Krishnakumar [37]	2023	CNN + LSTM	93.71
An <i>et al.</i> [5]	2024	EfficientNetB0 + DenseNet121 + Attention mechanism	95.15
Proposed	2025	Inception-ResNet + HFE + ECA	95.35

**TABLE 8.** Dataset summary for CT-scan, lung colon and mega datasets.

Dataset	Class	Train	Validation	Test	Total
CT-scan	Abnormal	465	59	261	785
	Normal	148	13	54	215
	Total	613	72	315	1000
Lung Colon	Colon_aca	3500	750	750	5000
	Colon_n	3500	750	750	5000
	Lung_aca	3500	750	750	5000
	Lung_n	3500	750	750	5000
	Lung_scc	3500	750	750	5000
	Total	17500	3,750	3,750	25000
Mega	Abnormal	465	–	261	726
	Colon_aca	4000	–	1,000	5000
	Lung_aca	4000	–	1,000	5000
	Lung_scc	4000	–	1,000	5000
	Normal	9489	–	2,288	11777
	Pneumonia	3875	–	390	4265
	Total	25829	–	5,939	31768

microscopic tissue characteristics. Finally, we created a Mega dataset combining all three imaging types to evaluate cross-modal generalization capabilities. This hierarchical validation strategy allows us to thoroughly examine the model performance across different imaging technologies and lung cancer manifestations, from radiological to pathological presentations.

Table 8 presents a comprehensive overview of the three datasets. The Chest CT-Scan dataset contains a variety of lung cancer subtypes, including large cell carcinoma (LCC), adenocarcinoma (AD) and squamous cell carcinoma (SCC) [39]. All of them are grouped under an abnormal class along with a normal category. This dataset consists of 1,000 CT images, partitioned into training, validation and testing subsets. Validating our model on this dataset is crucial, as CT imaging plays a pivotal role in lung cancer diagnosis owing to superior soft tissue contrast, which enables better tumor localization compared to X-rays. In other words, we employed this dataset to assess the developed model ability to accurately classify lung abnormalities in CT scans. While, the lung colon cancer

**FIGURE 7.** Generation of the mega dataset.

dataset [40] provides an additional layer of complexity, as it comprises histopathological images rather than radiological scans. This dataset includes 25,000 color images representing five distinct classes, lung adenocarcinoma (Lung\_aca), lung squamous cell carcinoma (Lung\_scc), benign lung tissue (Lung\_n), colon adenocarcinoma (Colon\_aca) and benign colon tissue (Colon\_n). Histopathological images offer a microscopic view of cellular structures, which differs significantly from the grayscale imaging of CT-scans and X-rays. Successfully adapting the proposed model to classify these images indicates the capacity to extract transferrable feature representations across different medical imaging modalities. This data set was randomly split into 17,500 training images, 3,750 validation images and 3,750 test images across the five classes.

To push the boundaries of our model generalization capabilities, we curated a new Mega dataset by using images from the X-ray, CT and histopathological datasets. This new data set consists of 31,768 images that span multiple modalities and cancer subtypes. Fig. 7 provides an illustration of how the mega dataset was constructed. The Mega dataset comprises abnormal cases from the CT dataset (726 images), lung and colon adenocarcinoma cases (10,000 images), lung squamous cell carcinoma cases (5,000 images), normal cases from X-ray, CT and histopathological datasets (11,777 images) as well as pneumonia cases from

**TABLE 9.** Classification results of the proposed model using CT-scan, lung colon and mega datasets.

Dataset	Validation/Testing	Class	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
CT-scan	Validation	Abnormal	100	100	100	100
		Normal	100	100	100	100
		Average	100	100	100	100
	Testing	Abnormal	99.68	99.62	100	99.81
		Normal	99.68	100	98.15	99.07
		Average	99.68	99.68	99.68	99.68
Lung Colon	Validation	Colon_aca	99.71	100	100	100
		Colon_n	99.71	100	100	100
		Lung_aca	99.71	100	98.53	99.26
		Lung_n	99.71	98.55	100	99.27
		Lung_scc	99.71	100	100	100
		Average	99.71	99.71	99.71	99.71
	Testing	Colon_aca	99.73	100	100	100
		Colon_n	99.73	100	100	100
		Lung_aca	99.73	100	98.80	99.40
		Lung_n	99.73	98.68	100	99.34
		Lung_scc	99.73	100	99.87	99.93
		Average	99.73	99.74	99.73	99.73
Mega	Testing	Abnormal	99.26	99.62	100	99.81
		Colon_aca	99.26	99.90	100	99.95
		Lung_aca	99.26	100	99.60	99.80
		Lung_scc	99.26	99.90	99.90	99.90
		Normal	99.26	99.60	98.56	99.08
		Pneumonia	99.26	92.31	98.46	95.29
		Average	99.26	99.29	99.26	99.27

the X-ray dataset (4,265 images). To ensure there was no data leakage or overlap, we preserved the original training, validation, and testing splits as defined by the source datasets. These datasets originate from different imaging centers and patient cohorts and were independently curated, eliminating the risk of shared instances or identifiers. The creation of this dataset addresses the heterogeneity in cancer detection, which is a paramount challenge in medical research. This diversity strengthens the model adaptability and ensures that it can function effectively across different clinical settings where a patient diagnosis might rely on multiple imaging techniques. Moreover, this dataset mitigates the overfitting risks associated with single dataset training by exposing the model to variations in imaging resolution, contrast and texture that naturally exist between different scanning technologies. It also enables multi-modal learning, which allows the model to develop a more holistic understanding of lung cancer visual characteristics across macroscopic and microscopic scales. Furthermore, the mega dataset facilitates the development of multi-source computer aided systems capable of assisting radiologists and pathologists in lung cancer detection.

Table 9 summarizes the performance of the proposed model across the various datasets. Besides, confusion metrics for the proposed model performance are displayed in

Fig. 8. For the CT-scan dataset, the model achieved perfect metrics during validation, with an accuracy of 100% for both abnormal and normal classes. During testing, the performance remained robust with an accuracy of 99.68%. Precision and recall were 99.62% and 100% for abnormal cases, while normal cases achieved 100% precision and 98.15% recall, resulting in F1-score of 99.81% and 99.07%, respectively. On the Lung Colon dataset, the model exhibited outstanding results across all classes. During testing, the average accuracy, precision, recall and F1-score were consistently 99.71%. Notably, Colon\_aca, Colon\_n, and Lung\_scc achieved perfect precision and recall, while Lung\_aca and Lung\_n showed slight variations, with precision and recall ranging between 98.53% and 100%. Validation results were similarly impressive, with an average accuracy of 99.73% and near-perfect precision and recall across all classes. For the Mega dataset, which includes multiple disease categories, the model maintained high performance with an accuracy of 99.26%. Precision and recall were particularly strong for Colon\_aca (99.90% and 100%) and Lung\_scc (99.90% and 99.90%). Abnormal cases achieved 99.62% precision and 100% recall, while normal cases and pneumonia showed slightly lower precision at 99.60% and 92.31%, respectively. Despite these variations, the overall F1-score remained high at 99.27%, revealing the model robustness across

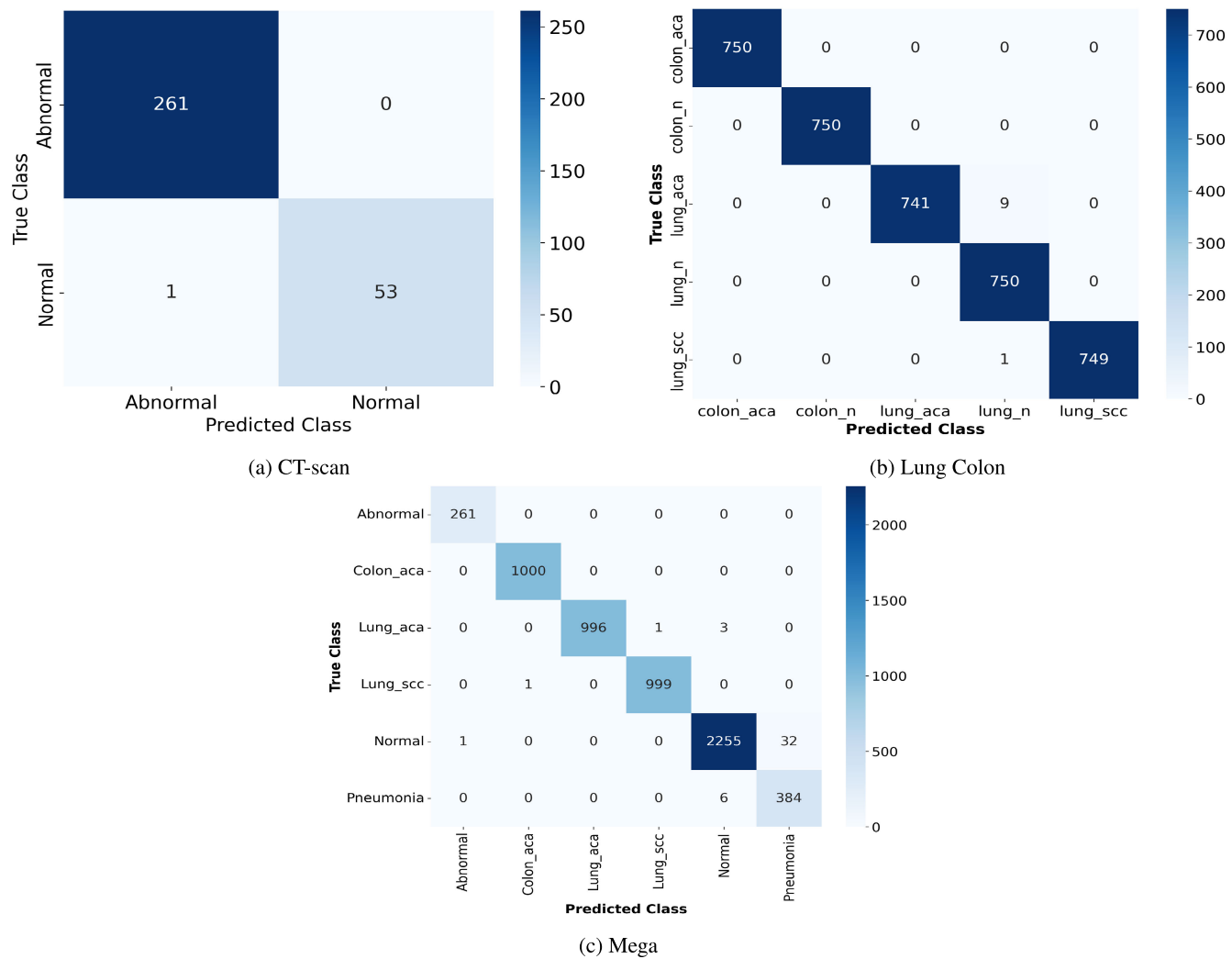


FIGURE 8. Confusion matrix for the three datasets.

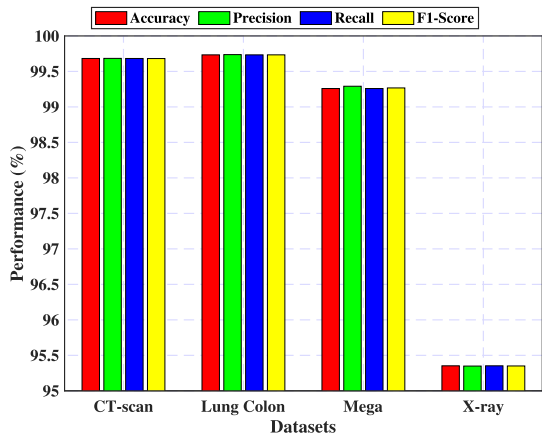


FIGURE 9. Performance of the proposed method across the four datasets.

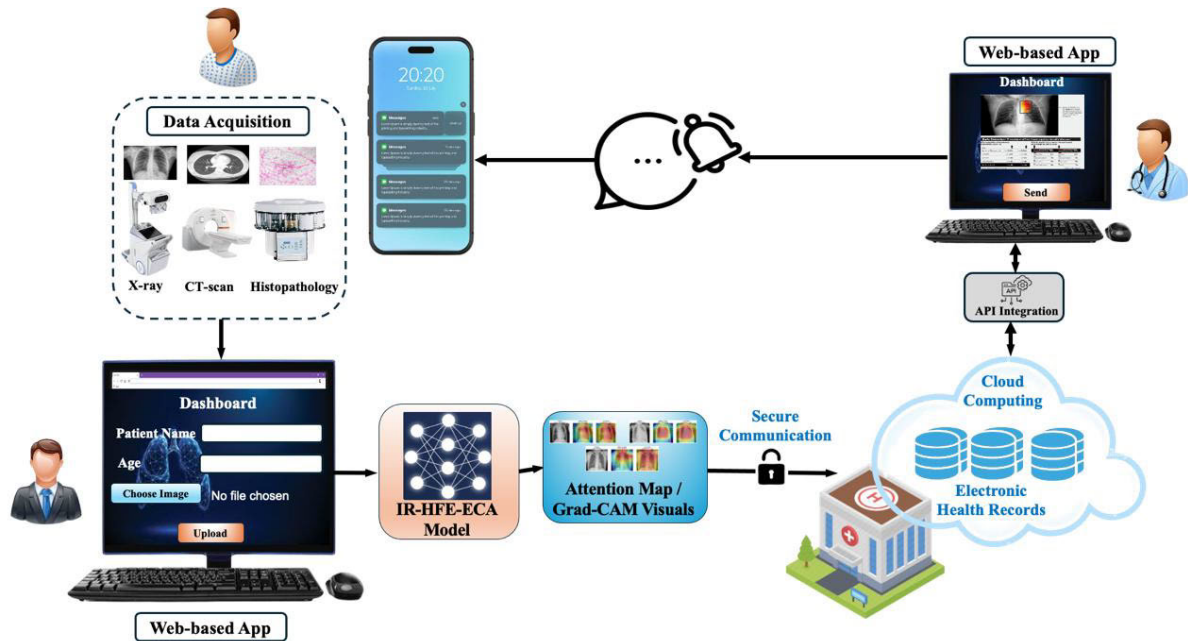
diverse datasets and disease categories. Our comprehensive experimental evaluation across the CT-scan, Lung Colon,

and Mega datasets demonstrates the model robust diagnostic capabilities. As evidenced in Fig. 9, the proposed architecture consistently achieved high performance across the four benchmark datasets. The ablation study not only validates the model superior lung cancer detection accuracy but also highlights its remarkable generalization across diverse imaging modalities and challenging clinical scenarios. These findings position our solution as a reliable diagnostic tool capable of handling medical imaging variability.

#### F. IMPLICATIONS

The findings of this study have significant implications for both clinical practice and future research in the domain of automated lung cancer diagnosis. From a clinical perspective, the ability of the proposed model to reliably classify different imaging modalities, while delivering dual explainability outputs (i.e., attention-based heatmaps and Grad-CAM overlays), makes it a practical tool for artificial intelligence (AI)-assisted diagnosis. Besides, in time-sensitive and





**FIGURE 10.** End-to-end multi-modal AI system for lung cancer screening and EHR integration.

equivocal cases, this visual explanations can reduce diagnostic ambiguity and bolster clinical decision-making confidence. Figure 10 illustrates the conceptual workflow for integrating the proposed model into an AI-driven clinical decision support system. The pipeline begins with data acquisition through a secure web portal where radiology staff upload images including chest X-rays, CT scans, and histopathology slides alongside basic patient metadata. This unified interface standardizes inputs across imaging modalities while eliminating the need for specialized software. Upon submission, the integrated AI inference engine automatically processes images using the proposed model architecture to generate both diagnostic classifications and visual interpretability outputs. The system produces predictions accompanied by attention heat maps that highlight salient features and Grad-CAM overlays that localize clinically relevant regions. All model outputs are securely transmitted through encrypted channels to a cloud-based electronic health records (EHR) repository, where they are recorded as timestamped entries via standardized API integration. This tight coupling with hospital information systems ensures regulatory compliance while maintaining complete audit trails of automated decisions within each patient medical record. For clinical deployment, the system offers physicians two complementary interfaces: a comprehensive diagnostic dashboard that presents AI-generated results alongside the original medical images for direct comparison, and a customizable notification system that delivers prioritized alerts through multiple communication channels when high-risk findings are identified. These interfaces collectively reduce workflow friction and accelerate clinical decision-making. The architecture demonstrates additional advantages through scalable cloud-based implementation, which supports both

batch processing of historical studies and on-demand expansion during peak utilization periods. Technically, the framework advances the field through the novel integration of Inception-ResNet feature extraction, hierarchical processing, and efficient channel attention mechanisms within a unified multimodal architecture, establishing new standards for interpretable AI systems in medical imaging.

#### IV. CONCLUSION

Lung cancer is a leading cause of mortality worldwide which necessitates innovative approaches for early and accurate diagnosis. Traditional diagnostic methods hang on manual interpretation of medical images. Nevertheless, these techniques are not only labor intensive but also prone to human error. Therefore, automated and scalable computer aided solutions are necessary. In this study, we introduce a novel DL framework that fuses multiple imaging modalities to address the limitations of existing models, such as single-modality reliance and suboptimal feature extraction. By combining the strengths of Inception-ResNet architectures, HFE blocks and ECA mechanisms, the proposed framework has the ability to dynamically recalibrate feature importance and capture multi-scale spatial features. This enabled the developed model to focus on highly informative regions within the medical images, which significantly increased diagnostic precision. Furthermore, the elimination of manual segmentation and extensive preprocessing streamlines the diagnostic process, making it more accessible for real-world clinical applications. The proposed model achieved the state-of-the-art performance in lung cancer detection, with accuracy rates of 95.35%, 99.68%, 99.73% and 99.26% on X-ray, CT, histopathological and mega datasets, respectively. Additionally, comparative experiments against established

CNN models and existing studies demonstrated the superior performance of our approach. Additionally, ablation studies validated the contribution of each component to the model overall effectiveness. Beyond lung cancer detection, the proposed model adaptability to multiple imaging modalities hold promise for extending application to other diseases. By exploiting advanced attention mechanisms and multi-modal data integration, our framework lays the groundwork for future AI-driven oncology research.

Although the results are encouraging, there is considerable potential for further refinement, particularly in optimizing the framework for real-time clinical applications. Therefore, future work will focus on exploring advanced attention mechanisms to boost the model performance. Furthermore, we plan to expand our evaluation using larger and more diverse datasets that encompass a wider range of medical conditions. Moreover, we will investigate the integration of federated learning to enable collaborative model training across distributed medical centers without compromising patient data privacy. In addition, optimizing the model for edge deployment will allow for low-latency, resource-efficient inference in decentralized or point-of-care environments, further increasing the practical applicability of the proposed framework.

## DECLARATIONS

### COMPETING INTERESTS

The authors declare no conflict of interest.

### CODE AVAILABILITY

The source code used in this study will be available upon reasonable request from the corresponding author.

### DATA AVAILABILITY

The datasets used in this study were downloaded from:

<https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia>

<https://www.kaggle.com/datasets/mohamedhanyyy/chest-ctscan-images>

<https://www.kaggle.com/datasets/andrewmvd/lung-and-colon-cancer-histopathological-images>

## REFERENCES

- [1] J. Maher, "Chimeric antigen receptor (CAR) T-Cell therapy for patients with lung cancer: Current perspectives," *OncoTargets Therapy*, vol. 16, pp. 515–532, Jul. 2023.
- [2] S. Maurya, S. Tiwari, M. C. Mothukuri, C. M. Tangeda, R. N. S. Nandigam, and D. C. Addagiri, "A review on recent developments in cancer detection using machine learning and deep learning models," *Biomed. Signal Process. Control*, vol. 80, Feb. 2023, Art. no. 104398.
- [3] A. Maged, M. Zhu, W. Gao, and M. Hosny, "Lightweight deep learning model for automated STN localization using MER in Parkinson's disease," *Biomed. Signal Process. Control*, vol. 96, Oct. 2024, Art. no. 106640.
- [4] M. Hosny, A. M. Elshenhab, and A. Maged, "Explainable AI-based method for brain abnormality diagnostics using MRI," *Biomed. Signal Process. Control*, vol. 100, Feb. 2025, Art. no. 107184.
- [5] Q. An, W. Chen, and W. Shao, "A deep convolutional neural network for pneumonia detection in X-ray images with attention ensemble," *Diagnostics*, vol. 14, no. 4, p. 390, Feb. 2024.
- [6] H. Bhatt and M. Shah, "A convolutional neural network ensemble model for pneumonia detection using chest X-ray images," *Healthcare Anal.*, vol. 3, Nov. 2023, Art. no. 100176.
- [7] G. Mohandass, G. Hari Krishnan, D. Selvaraj, and C. Sridhathan, "Lung cancer classification using optimized attention-based convolutional neural network with DenseNet-201 transfer learning model on CT image," *Biomed. Signal Process. Control*, vol. 95, Sep. 2024, Art. no. 106330.
- [8] M. Hammad, M. ElAffendi, M. Asim, A. A. Abd El-Latif, and R. Hashiesh, "Automated lung cancer detection using novel genetic TPOT feature optimization with deep learning techniques," *Results Eng.*, vol. 24, Dec. 2024, Art. no. 103448.
- [9] G. M. M. Alshmrani, Q. Ni, R. Jiang, H. Pervaiz, and N. M. Elshennawy, "A deep learning architecture for multi-class lung diseases classification using chest X-ray (CXR) images," *Alexandria Eng. J.*, vol. 64, pp. 923–935, Nov. 2022.
- [10] S. Bharati, P. Podder, and M. R. H. Mondal, "Hybrid deep learning for detecting lung diseases from X-ray images," *Informat. Med. Unlocked*, vol. 20, Jan. 2020, Art. no. 100391.
- [11] A. Bhandary, G. A. Prabhu, V. Rajinikanth, K. P. Thanaraj, S. C. Satapathy, D. E. Robbins, C. Shasky, Y.-D. Zhang, J. M. R. S. Tavares, and N. S. M. Raja, "Deep-learning framework to detect lung abnormality—A study with chest X-ray and lung CT scan images," *Pattern Recognit. Lett.*, vol. 129, pp. 271–278, Jan. 2020.
- [12] L. S. K., S. N. Mohanty, S. K., A. N., and G. Ramirez, "Optimal deep learning model for classification of lung cancer on CT images," *Future Gener. Comput. Syst.*, vol. 92, pp. 374–382, Mar. 2019.
- [13] A. A. Shah, H. A. M. Malik, A. Muhammad, A. Alourani, and Z. A. Butt, "Deep learning ensemble 2D CNN approach towards the detection of lung cancer," *Sci. Rep.*, vol. 13, no. 1, p. 2987, Feb. 2023.
- [14] V. Rajasekar, M. Vaishnav, S. Premkumar, V. Sarveshwaran, and V. Rangaraaj, "Lung cancer disease prediction with ct scan and histopathological images feature analysis using deep learning techniques," *Results Eng.*, vol. 18, May 2023, Art. no. 101111.
- [15] M. Masud, N. Sikder, A.-A. Nahid, A. K. Bairagi, and M. A. AlZain, "A machine learning approach to diagnosing lung and colon cancer using a deep learning-based classification framework," *Sensors*, vol. 21, no. 3, p. 748, Jan. 2021.
- [16] M. Toğaçar, "Disease type detection in lung and colon cancer images using the complement approach of inefficient sets," *Comput. Biol. Med.*, vol. 137, Oct. 2021, Art. no. 104827.
- [17] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Y. Ng, "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning," 2017, *arXiv:1711.05225*.
- [18] A. Rehman, M. Harouni, F. Zogh, T. Saba, M. Karimi, F. S. Alamri, and G. Jeon, "Detection of lungs tumors in CT scan images using convolutional neural networks," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 21, no. 4, pp. 769–777, Jul. 2024.
- [19] E. Ayan, B. Karabulut, and H. M. Ünver, "Diagnosis of pediatric pneumonia with ensemble of deep convolutional neural networks in chest X-ray images," *Arabian J. Sci. Eng.*, vol. 47, no. 2, pp. 2123–2139, Feb. 2022.
- [20] S. Goyal and R. Singh, "Detection and classification of lung diseases for pneumonia and COVID-19 using machine and deep learning techniques," *J. Ambient Intell. Humanized Comput.*, vol. 14, no. 4, pp. 3239–3259, Apr. 2023.
- [21] P. Xie, X. Zhao, and X. He, "Improve the performance of CT-based pneumonia classification via source data reweighting," *Sci. Rep.*, vol. 13, no. 1, p. 9401, Jun. 2023.
- [22] F. Wang, X. Li, R. Wen, H. Luo, D. Liu, S. Qi, Y. Jing, P. Wang, G. Deng, C. Huang, T. Du, L. Wang, H. Liang, J. Wang, and C. Liu, "Pneumonia-plus: A deep learning model for the classification of bacterial, fungal, and viral pneumonia based on CT tomography," *Eur. Radiol.*, vol. 33, no. 12, pp. 8869–8878, Jun. 2023.
- [23] N. Kumar, M. Sharma, V. P. Singh, C. Madan, and S. Mehandia, "An empirical study of handcrafted and dense feature extraction techniques for lung and colon cancer classification from histopathological images," *Biomed. Signal Process. Control*, vol. 75, May 2022, Art. no. 103596.
- [24] M. A. Talukder, M. M. Islam, M. A. Uddin, A. Akhter, K. F. Hasan, and M. A. Moni, "Machine learning-based lung and colon cancer detection using deep feature extraction and ensemble learning," *Expert Syst. Appl.*, vol. 205, Nov. 2022, Art. no. 117695.

- [25] S. Mehmood, T. M. Ghazal, M. A. Khan, M. Zubair, M. T. Naseem, T. Faiz, and M. Ahmad, "Malignancy detection in lung and colon histopathology images using transfer learning with class selective image processing," *IEEE Access*, vol. 10, pp. 25657–25668, 2022.
- [26] G. Brauwiers and F. Frasincar, "A general survey on attention mechanisms in deep learning," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 4, pp. 3279–3298, Apr. 2023.
- [27] S.-M. Cha, S.-S. Lee, and B. Ko, "Attention-based transfer learning for efficient pneumonia detection in chest X-ray images," *Appl. Sci.*, vol. 11, no. 3, p. 1242, Jan. 2021.
- [28] K. Wang, P. Jiang, J. Meng, and X. Jiang, "Attention-based DenseNet for pneumonia classification," *IRBM*, vol. 43, no. 5, pp. 479–485, Oct. 2022.
- [29] P. Mooney. (2018). *Chest X-ray Images (Pneumonia)*. Accessed: Sep. 1, 2024. [Online]. Available: <https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia>
- [30] M. Q. R. Pembury Smith and G. D. Ruxton, "Effective use of the McNemar test," *Behav. Ecol. Sociobiol.*, vol. 74, no. 11, pp. 1–9, Nov. 2020.
- [31] D. C. Hoaglin, "Misunderstandings about q and 'cochran's q test' in meta-analysis," *Statist. Med.*, vol. 35, no. 4, pp. 485–495, 2016.
- [32] S. Sharma and K. Guleria, "A deep learning based model for the detection of pneumonia from chest X-ray images using VGG-16 and neural networks," *Proc. Comput. Sci.*, vol. 218, pp. 357–366, Jan. 2023.
- [33] O. Stephen, M. Sain, U. J. Maduh, and D.-U. Jeong, "An efficient deep learning approach to pneumonia classification in healthcare," *J. Healthcare Eng.*, vol. 2019, Mar. 2019, Art. no. 4180949.
- [34] G. Liang and L. Zheng, "A transfer learning method with deep residual network for pediatric pneumonia diagnosis," *Comput. Methods Programs Biomed.*, vol. 187, Apr. 2020, Art. no. 104964.
- [35] M. Salehi, R. Mohammadi, H. Ghaffari, N. Sadighi, and R. Reiazi, "Automated detection of pneumonia cases using deep transfer learning with paediatric chest X-ray images," *Brit. J. Radiol.*, vol. 94, no. 1121, May 2021, Art. no. 20201263.
- [36] A. Mabrouk, R. P. D. Redondo, A. Dahou, M. Abd Elaziz, and M. Kayed, "Pneumonia detection on chest X-ray images using ensemble of deep convolutional neural networks," *Appl. Sci.*, vol. 12, no. 13, p. 6448, Jun. 2022.
- [37] V. J. M. Jaya and S. Krishnakumar, "Multi-classification approach for lung nodule detection and classification with proposed texture feature in X-ray images," *Multimedia Tools Appl.*, vol. 83, no. 2, pp. 3497–3524, Jan. 2024.
- [38] I. A. Elgendy, M. Hosny, M. Albashrawi, and S. Alsenan, "Dual-stage explainable ensemble learning model for diabetes diagnosis," *Expert Syst. Appl.*, vol. 274, Feb. 2025, Art. no. 126899.
- [39] M. Hany. (2020). *Chest Ct-Scan Images Dataset*. Accessed: Sep. 1, 2024. [Online]. Available: <https://www.kaggle.com/datasets/mohamedhanyyy/chest-ctscan-images>
- [40] Larxel. (2020). *Lung and Colon Cancer Histopathological Images*. Accessed: Sep. 1, 2024. [Online]. Available: <https://www.kaggle.com/datasets/andrewmvd/lung-and-colon-cancer-histopathological-images>



**MOHAMED HOSNY** received the M.Sc. degree in biomedical engineering from the Biomedical Engineering Department, Benha Faculty of Engineering, Benha University, Egypt, in 2017, and the Ph.D. degree in biomedical engineering from the School of Life Science and Technology, Harbin Institute of Technology, Harbin, China, in 2022. He is currently with the King Fahd University of Petroleum and Minerals, Dammam, Saudi Arabia, as a Postdoctoral Fellow. He is also an Assistant

Professor with Benha Faculty of Engineering, Benha University. His research interests include computer vision, machine learning, and image and signal processing. Furthermore, he has published 19 papers in prestigious journals and conferences, including biomedical signal processing and control and biocybernetics and biomedical engineering.



**IBRAHIM A. ELGENDY** received the B.Sc. and M.Sc. degrees from the Computer Science Department, Faculty of Computers and Information, Menoufia University, Egypt, in 2010 and 2016, respectively, and the Ph.D. degree from the School of Computer Science and Technology, Harbin Institute of Technology, China, in 2021. He is currently with the King Fahd University of Petroleum and Minerals, Dammam, Saudi Arabia, as a Postdoctoral Fellow. He is also an Assistant

Professor with the Computer Science Department, Faculty of Computers and Information, Menoufia University. His research interests include cloud computing, mobile edge computing, the Internet of Things, security, and distributed computing. Furthermore, he has published over 40 papers in prestigious journals and conferences, including IEEE INTERNET OF THINGS JOURNAL (IEEE IoT), IEEE TRANSACTIONS ON CLOUD COMPUTING (IEEE TCC), IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS (IEEE TII), IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT, FGCS, and IEEE ACCESS. He served as an Associate and Guest Editor for many international journals, such as *Sensors* (MDPI), the *International Journal of Information Security and Privacy*, and the *International Journal of Smart Security Technologies*; and a Reviewer of more than 130 articles for many prestigious journals, such as IEEE TCC, IEEE IoT, IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE TRANSACTIONS ON MOBILE COMPUTING, IEEE TII, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE ACCESS, *PLOS One*, *Sensors*.



**MOUSA AHMAD ALBASHRAWI** received the bachelor's degree in management information systems from the King Fahd University of Petroleum and Minerals (KFUPM), the master's degree in information systems (health information technology management) from the University of Colorado Denver, and the Ph.D. degree in business administration (management information systems) from the University of Massachusetts Lowell. He is currently an Associate Professor with KFUPM,

specializing in FinTech, AI in healthcare, and business analytics. He has contributed to advancing FinTech research and digital economy initiatives at KFUPM. He is also the Director of the Interdisciplinary Research Center for Finance and Digital Economy, KFUPM, and also acts as the Coordinator of the Master's Program in Business Analytics. Additionally, he co-founded Saudi Chapter of the Association for Information Systems (AIS) and has received numerous awards recognizing his contributions to academic mentoring and leadership. His research interests include data analytics, AI and machine learning, FinTech, and social media. He has published extensively in leading academic journals and has received multiple best paper awards. He also serves as an Associate Editor for Communications of the AIS (CAIS) and an Editorial Board Member for *Arab Journal of Administrative Sciences*.

...