IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 20 January 2025, accepted 2 February 2025, date of publication 4 February 2025, date of current version 12 February 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3539122

RESEARCH ARTICLE

Attention Enhanced InceptionNeXt-Based Hybrid Deep Learning Model for Lung Cancer Detection

BURHANETTIN OZDEMIR[1], EMRAH ASLAN[2], AND ISHAK PACAL[3,4]

[1]Department of Operations and Project Management, College of Business, Alfaisal University, Riyadh 11533, Saudi Arabia
[2]Department of Computer Engineering, Faculty of Engineering and Architecture, Mardin Artuklu University, 47000 Mardin, Türkiye
[3]Department of Computer Engineering, Faculty of Engineering, Igdir University, 76000 Iğdır, Türkiye
[4]Department of Electronics and Information Technologies, Faculty of Architecture and Engineering, Nakhchivan State University, AZ 7012 Nakhchivan, Azerbaijan

Corresponding author: Burhanettin Ozdemir (bozdemir@alfaisal.edu)

This work was funded by Alfaisal University, which funds research initiatives aimed at advancing knowledge and innovation in alignment with its commitment to academic excellence.

ABSTRACT Lung cancer is the most common cause of cancer-related mortality globally. Early diagnosis of this highly fatal and prevalent disease can significantly improve survival rates and prevent its progression. Computed tomography (CT) is the gold standard imaging modality for lung cancer diagnosis, offering critical insights into the assessment of lung nodules. We present a hybrid deep learning model that integrates Convolutional Neural Networks (CNNs) with Vision Transformers (ViTs). By optimizing and integrating grid and block attention mechanisms with InceptionNeXt blocks, the proposed model effectively captures both fine-grained and large-scale features in CT images. This comprehensive approach enables the model not only to differentiate between malignant and benign nodules but also to identify specific cancer subtypes such as adenocarcinoma, large cell carcinoma, and squamous cell carcinoma. The use of InceptionNeXt blocks facilitates multi-scale feature processing, making the model particularly effective for complex and diverse lung nodule patterns. Similarly, including grid attention improves the model's capacity to identify spatial relationships across different sections of the picture, whereas block attention focuses on capturing hierarchical and contextual information, allowing for precise identification and categorization of lung nodules. To ensure robustness and generalizability, the model was trained and validated using two public datasets, Chest CT and IQ-OTH/NCCD, employing transfer learning and pre-processing techniques to improve detection accuracy. The proposed model achieved an impressive accuracy of 99.54% on the IQ-OTH/NCCD dataset and 98.41% on the Chest CT dataset, outperforming state-of-the-art CNN-based and ViT-based methods. With only 18.1 million parameters, the model provides a lightweight yet powerful solution for early lung cancer detection, potentially improving clinical outcomes and increasing patient survival rates.

INDEX TERMS CNN, deep learning, lung cancer, ViT.

I. INTRODUCTION

Lung cancer is one of the most common and dangerous forms of cancer globally [1]. Known as the leading cause of cancer-related deaths, lung cancer is often fatal when diagnosed late. This disease is caused by the formation of malignant tumors that start in the respiratory tract and usually spread rapidly [2]. It affects millions of people each year and is common in both men and women. By 2024, there will likely be 611,720 cancer-related deaths and 2,001,140 new instances of cancer in the US alone. According to 2024 statistics, 234,580 lung cancer cases have been identified. 125,070 deaths from lung cancer were recorded [3]. Lung cancer is distinguished by uncontrolled cell proliferation, which affects normal pulmonary function and accelerates the disease.

Lung cancer is classified into two types: small-cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). More than 85% of all cases of lung cancer are non-small cell lung cancer, making it the most common type. SCLC tends to be a more aggressive and rapidly spreading type of cancer [4].

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

Lung cancer is often diagnosed at an advanced stage, which limits treatment options and lowers survival rates. Therefore, early diagnosis of lung cancer can significantly increase the patient's survival rate and positively impact the treatment process [5].

Early detection is crucial for treating lung cancer. Early detection of lung cancer can greatly extend a patient's life expectancy and improve response to therapy. Traditionally, lung cancer has been diagnosed using X-rays, MRI, PET, and biopsy. However, these methods can be time-consuming, expensive, and invasive. Detecting early-stage lung cancer is particularly challenging due to the difficulty of identifying small nodules. CT scans provide important information in the evaluation of lung nodules and improve the diagnostic process for specialists. However, manual review of CT scans is a tedious process prone to human error [6]. These methods are used to determine the presence and stage of cancer; however, each of these techniques has certain limitations. X-rays and CT scans can be effective in detecting lung nodules, but these imaging modalities are often insufficient to make a definitive diagnosis of cancer [7]. Biopsy and bronchoscopy are more invasive procedures and may pose additional risks to the patient [8]. Traditional methods are often costly, and time-consuming and can negatively impact patient comfort. Therefore, there is a growing need for faster, cheaper, and less invasive diagnostic methods [9].

In recent years, the medical sector has experienced a significant increase in the utilization of artificial intelligence (AI) and deep learning (DL) [10]. In particular, DL algorithms show superior performance in identifying complex patterns and anomalies by training on large datasets [11]. CNNs have found significant success, particularly in image processing and medical image analysis [12]. AI-based approaches have significant potential in early diagnosis and classification of complex diseases such as lung cancer [13]. These techniques can speed up the diagnosis process and improve its accuracy by automatically analyzing medical images [14], [15].

DL has fundamentally transformed medical image analysis, introducing powerful tools for diagnosing and detecting a wide range of diseases. By utilizing large datasets and advanced algorithms, these techniques can reveal intricate patterns and features in medical images that are typically beyond the capabilities of human interpretation. This automation has brought about substantial improvements in diagnostic accuracy, reduced the time needed for analysis, and minimized the possibility of errors. Among its diverse applications, the early detection of cancer using medical imaging has gained significant attention, offering the potential to greatly enhance survival rates and patient outcomes. Modalities such as CT, magnetic resonance imaging (MRI), and histopathology have become key areas where DL methods have been applied with notable success.

CNNs stand out as some of the most effective DL techniques in medical imaging, particularly for early cancer diagnosis. CNNs can efficiently process and analyze vast amounts of image data, making them well-suited for identifying cancerous tissues, cells, or lesions with precision. Their versatility has been demonstrated in tasks such as detecting abnormalities in CT scans and classifying lung cancer cells, where they have consistently delivered high performance.

Building on the capabilities of CNNs, ViTs have emerged as a next-generation approach, offering groundbreaking advancements in medical imaging. Unlike traditional CNNs, ViTs use attention mechanisms to extract features from image data, enabling them to capture both local details and global patterns with remarkable efficiency. These models have shown exceptional promise in analyzing complex datasets and detecting nuanced features in medical images. However, while ViTs excel in handling large-scale image data, CNNs often retain an edge in tasks that require intricate feature extraction, such as the detection and classification of lung cancer lesions. The early adoption of these DL models in diagnostic workflows has the potential to streamline the detection of lung cancer, leading to more accurate treatment strategies and improved outcomes for patients.

This study addresses significant challenges in lung cancer diagnosis, focusing on the limitations of current CT-based methods. Existing AI and DL approaches for CT imaging often struggle to accurately capture both fine-grained and large-scale features in medical images. These limitations can result in suboptimal diagnostic accuracy, which may delay treatment. Additionally, manual reviews of CT scans remain time-intensive and prone to human error. To overcome these challenges, this research introduces a hybrid DL model that integrates grid and block attention mechanisms with InceptionNeXt blocks. This novel architecture is designed to enhance feature extraction and improve the accuracy of lung cancer detection and classification.

Training and evaluation of the model were performed using two publicly available datasets with limited demographic diversity, IQ-OTH/NCCD, and Chest CT. This may partly limit the generalizability of the model in different populations and geographical regions. Furthermore, an approach based on CT images only was adopted, and future studies aim to improve model performance by integrating multimodal data such as biomarkers or patient history.

The study is motivated by the pressing need for faster and more precise diagnostic tools to improve outcomes for lung cancer patients. CT imaging of lung nodules presents unique challenges due to the subtle and diverse appearance of these abnormalities. This study validates the proposed hybrid model using two public datasets, Chest CT and IQ-OTH/NCCD, which include comprehensive annotations of cancerous and non-cancerous nodules. Pre-processing techniques, such as contrast enhancement and noise reduction, are employed to improve image quality and ensure robust detection accuracy. These measures support the rigorous evaluation of the model's performance in realistic clinical scenarios.

The contributions of this paper to the literature are as follows:

- A novel hybrid DL model is presented, combining CNNs and ViTs to address critical challenges in lung cancer detection.
- The model demonstrates superior performance compared to five leading CNN-based and five leading ViT-based models trained under identical conditions, achieving state-of-the-art accuracy of 99.54% on the IQ-OTH/NCCD dataset and 98.41% on the Chest CT dataset.
- Integration of advanced grid and block attention mechanisms with InceptionNeXt blocks enhances the model's ability to capture multi-scale features, spatial relationships, and contextual information.
- With a lightweight design of only 18.1 million parameters, the model offers an efficient and scalable solution for early lung cancer detection, ensuring practical applicability in clinical environments.

II. LITERATURE REVIEW

Early detection and treatment of lung cancer is critical to saving lives. Lung cancer is a life-threatening condition resulting from the uncontrolled growth of malignant cells in one or both lungs, which, if not treated early, can spread to other organs. As such, there is a significant need for an effective computer-aided diagnosis (CAD) system capable of detecting and classifying lung cancer with greater accuracy. In this section, the methods, techniques, approaches, and stages of lung image processing for lung cancer detection by various authors in the literature will be comprehensively discussed.

Sabzalian et al. study aims to accurately classify malignant and benign tumors using an improved bidirectional recurrent neural network (RNN) and Ebola optimization search algorithm, a new method for early detection of lung cancer. The study on the IQ-OTH/NCCD lung cancer dataset showed superior performance [16]. Ma et al. presented GoogLeNet-AL, a novel CNN architecture for lung cancer detection. GoogLeNet-AL integrated several innovative components to efficiently capture multi-scale features and was implemented on the PyTorch platform, demonstrating superior performance with higher F1 score and accuracy. The model outperforms traditional GoogLeNet and other baseline models, providing an important tool for lung nodule detection and classification [17].

Gautam et al. propose a new DL ensemble model for lung cancer detection consisting of ResNet-152, DenseNet 169 and EfficientNet-B7 models. The model is weight-optimized using ROC-AUC and F1 scores. It showed superior performance on the LIDC-IDRI dataset with 97.23% accuracy and 98.6% sensitivity. This method significantly reduces the number of false negatives [18]. Wani et al. used DeepXplainer, a novel explicable DL technique, for lung cancer detection. This ConvXGB-based model learns features with DL, performs classification with XGBoost, and then provides explanations of the predictions with the SHAP method. The method was tested on the Survey Lung Cancer dataset and showed superior performance with 97.43% accuracy, 98.71% sensitivity, and 98.08% F1 score [19].

Heidari introduces a blockchain-based federated learning (FL) method for lung cancer detection using CT scans. This method uses blockchain technology to train a global DL model, while maintaining the privacy of data collected from different hospitals. Experiments have shown that the method is effective with high accuracy on Cancer Imaging Archive, Kaggle Data Science Bowl, LUNA 16, and local datasets [20]. Bushara et al. used an innovative capsule network combination called VGG-CapsNet for lung cancer detection and classification. VGG-CapsNet overcame the shortcomings of CNNs in recognizing fine-grained spatial relationships and demonstrated superior performance on LIDC-IDRI and Kaggle datasets with high accuracy, sensitivity, and specificity [21]. In Raza's study, they used Lung-EffNet, a new model based on EfficientNet, to classify lung cancer using CT scans. Lung-EffNet was developed based on the architecture of EfficientNet and improved by adding classification layers. The model achieved superior performance with high accuracy on the IQ-OTH/NCCD dataset, with faster and more efficient results compared to other CNN models [22].

Subash and Kalaivani presented a two-stage classification model for lung cancer detection and staging using advanced DL techniques. The model uses a modified version of U-Net with dual attention and pyramid atrous pooling and a combination of Xception and custom CNN to discriminate between abnormal and normal cases and improve target segmentation accuracy. In the second stage, additional spatial features are extracted from abnormal features and lung cancer staging is performed with a hybrid adaptive learning neural network and the performance of the model has shown successful results on LIDC-IDRI and NSCLC datasets [23].

Nahiduzzaman et al. present an innovative approach combining LPDCNN and Ridge-ELM models for the classification of three types of lung cancer and normal lung tissue. Image quality is improved by using CLAHE and Gaussian blur, and LPDCNN extracts discriminative features with low computational cost. The model, whose tractability is improved by SHAP integration, achieves high accuracy and recall in four-class and binary classifications [24]. He et al. presented an automated Darknet-based immunohistochemistry (IHC) scoring system for IL-24 in lung cancer. Overcoming the challenges of complex backgrounds and overlapping cells in IHC images, the proposed system directly calculates clinical scores using a block attention mechanism and a Darknet-based scoring network. Experiments were conducted on 5000 manually anodized IHC images, and it was shown that the system can greatly support clinical diagnosis and treatment [25].

Gowthamy et al. created a unique hybrid model for rapid and accurate lung and colon cancer diagnosis by combining Kernel Extreme Learning Machine (KELM) with pre-trained DL models including ResNet-50, InceptionV3, and DenseNet. These pre-trained models offer a solid basis

for feature extraction by capturing complicated patterns typical of malignant tissue. KELM provides fast and accurate classification by efficiently processing the high-dimensional feature space. Optimizing the model parameters with the MB-DMOA algorithm provides better convergence to the global optimum by reducing the risk of getting stuck in the local optimum [26]. Tran et al. investigated how DL approaches can be applied to lung cancer decision-making and treatment development using omics data. This study summarizes the current state of DL-based lung cancer genomics research and discusses future research directions [27]. Lanjewar et al. identified four categories of lung cancer using a Kaggle dataset of chest CT images. They proposed a unique DL based method with modifications and additional layers on the DenseNet201 model. For the features obtained with this method, two feature selection methods were used and applied to different machine learning (ML) classifiers. The results of the study were evaluated with high accuracy up to perfection [28].

Naseer et al. employed a modified U-Net lobe segmentation and nodule detection model to accurately identify and classify lung cancer in CT data. The first stage uses modified U-Net to segment the lobes, while the second step extracts potential nodules. Finally, the nodules are classified as cancer or non-cancer using modified AlexNet and support vector machine. Experiments on the LUAN16 dataset show that the proposed methodology achieves 97.98% accuracy and 98.84% sensitivity for lung cancer classification, with promising results in other performance metrics [29].

Reference [30] proposed a complex and confusing classification model for diagnosing lung cancer that integrated multiple deep networks, such as BEiT, DenseNet, and Sequential CNN, using various ensembling techniques like AND, OR, Weighted Box Fusion, and Boosting. They verified the ensemble model on Chest CT-Scan Images Dataset for 98% accuracy, instead of the previous single-model-based approaches. These findings further underscore the efficacy of ensemble methods in enhancing diagnostic accuracy for medical imaging tasks [30].

Reference [31] proposed a three-stage lung cancer detection approach including image segmentation, fine-tuning of the weighted VGG deep network, and real-time inference using Nvidia Tensor-RT. Their system, evaluated on 19,419 CT lung tissue slices of the LIDC dataset, reported an accuracy of 93.2%, an F1 score of 0.93, and Cohen's kappa of 0.85, proving the feasibility for real-time diagnostics by [31]. Xiao et al. [32] proposed MFMANet: a multi-feature multi-attention network that improves the classification accuracy of NSCLC subtypes. This network is designed with two new modules: MSAM gives attention to multi-scale spatial channel information, and MFGLA pays attention to multi-feature fusion with global attention. It can enhance the detection of small lesion regions. Validated by two public datasets, MFMANet reports much better performance compared to existing models such as ResNet18 and ShuffleNetv2, hence proving its effectiveness in the classification of NSCLC subtypes. Moreover, Lin and Yang [33] presented the F-CFNN, a convolutional fuzzy neural network that uses two times convolutional layers along with two times pooling layers, for feature extraction, through which classification is based on the fuzzy neural network. Five feature fusion methods, such as GMP and Network Mapping Fusion, have been used. By allowing the Taguchi method to determine optimal parameter determination, the model achieved 99.98% rightness, showing a pretty improved performance in lung cancer classification compared with the works presented before.

Lakshmanaprabu et al. [34] proposed a novel framework using an Optimized Deep Neural Network coupled with Linear Discriminant Analysis for lung cancer diagnosis from CT images. The developed framework classified lung nodules as malignant or benign and utilized LDA to reduce the dimensionality of the extracted deep features. When combined with the optimization of ODNN by the Modified Group Search Optimization Algorithm, it resulted in a sensitivity of 96.2% and specificity of 94.2%, with an overall accuracy of 94.56%, proving the framework's efficiency in detecting lung cancer. Collectively, these advancements highlight how AI is revolutionizing lung cancer diagnosis through ensemble techniques, attention mechanisms, optimization strategies, and more, offering more accurate, interpretable, and real-time solutions for clinical applications.

Mahum and Al-Salman propose Lung-RetinaNet, a novel RetinaNet-based method for lung tumor detection. Using innovative techniques such as multi-scale feature fusion and context module, this methodology both augments semantic information from deep network layers and integrates contextual information to effectively identify lung tumors. The proposed methodology has shown superior performance compared to existing DL-based methods [35]. Atiya et al. explored the efficacy of DL approaches, especially transfer learning and deep convolutional neural networks (DCNN), in lung cancer categorization. The proposed dual-state transfer learning approach attempts to reliably diagnose lung cancer kinds from chest CT scan pictures using pre-trained models like ResNet50. The study shows that the proposed model can outperform existing techniques with high accuracy and classification performance [36].

In the literature, studies on the success of AI algorithms in lung cancer diagnosis have achieved high accuracy and sensitivity rates with different techniques. These studies with various AI and DL approaches have provided significant advances in lung cancer diagnosis and have shown promising results in the field of early detection and classification.

III. MATERIALS AND METHODS

This section presents an improved hybrid model of the MaxViT DL algorithm for lung cancer classification. Details of the publicly available datasets used for training and testing are described. The proposed method combines the power of a vision transducer with sophisticated data augmentation and transfer learning strategies to effectively detect and classify lung cancer with high sensitivity and specificity. To ensure

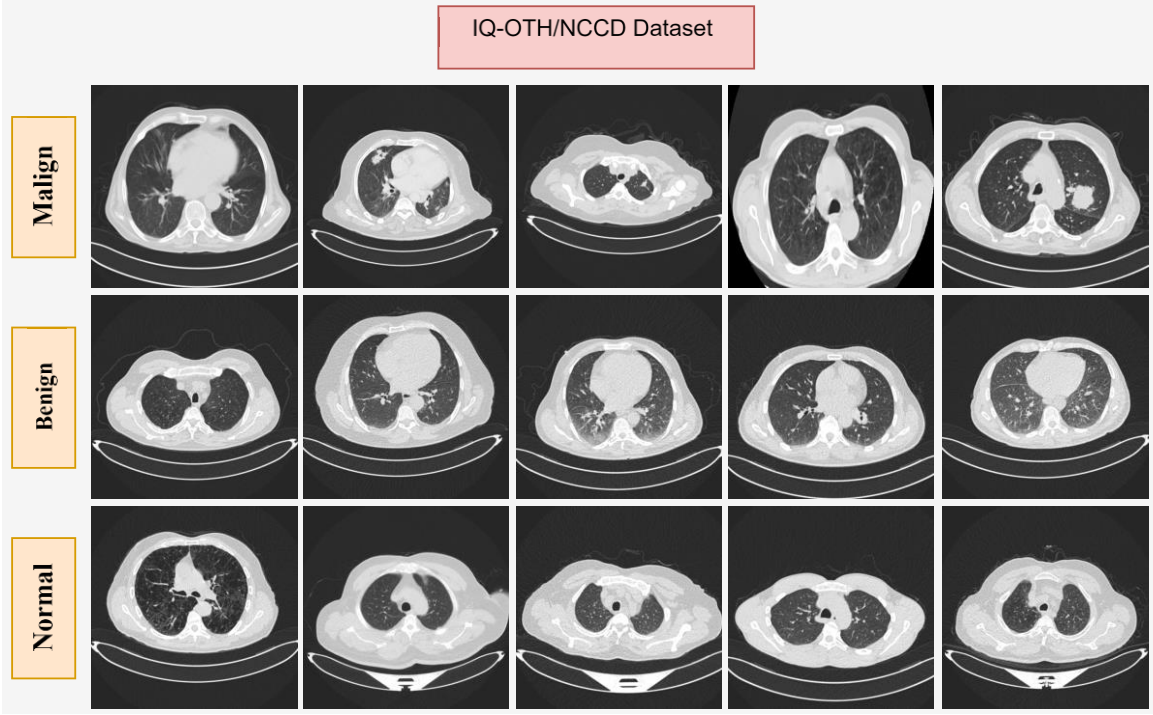

FIGURE 1. Sample images for each class from the IQ-OTH/NCCD dataset.

TABLE 1. Basic information about the IQ-OTH/NCCD dataset.

Class	Category	Number of Images
1	Malign	561
2	Benign	120
3	Normal	416
Total		1197

reproducibility and demonstrate accuracy, the studies were performed on two datasets.

A. DATASET

The success of DL algorithms has been driven by their ability to learn from large datasets. The quality and size of the dataset play a critical role in enabling the model to learn accurate and generalizable features. A large dataset reduces bias, prevents overfitting or underfitting, and ensures balanced performance across different subgroups. It also facilitates transfer learning, helping pre-trained models to adapt quickly to new tasks. However, high-quality datasets for lung cancer diagnosis remain limited in the literature. Among the datasets that have been shown to be effective in this area, the Chest CT and IQ-OTH/NCCD datasets stand out. Both datasets have been widely used in AI-based research for lung cancer diagnosis and have been endorsed by many experts. In this study, only CT images were used because CT is considered the most reliable imaging modality for early detection and subtype classification of lung nodules due to the high-resolution anatomical details it provides. Due to their open access and comprehensive nature, these datasets are an important resource for researchers and increase the effectiveness of models developed for lung cancer detection. The diversity of classifications between these two datasets allowed the model to perform successfully in complex tasks such as both general lung cancer detection and subtype classification. At the same time, the demographic differences strengthened the applicability of the model to a wide range of patients in the clinical setting. These two datasets used in our study allowed us to evaluate and validate the overall performance of our model due to their comprehensive, reliable and clinically rich content, strengthening the robustness of our results and their place in the literature.

The Chest CT and IQ-OTH/NCCD datasets were selected for their proven effectiveness in lung cancer diagnostic research, open access and comprehensive descriptions. These datasets support robust model development and evaluation processes by providing high quality imaging data with detailed descriptions for different subtypes. The IQ-OTH/NCCD dataset primarily represents a specific region, while the Chest CT dataset focuses on subtype-specific descriptions. Nevertheless, both datasets offer significant clinical value due to their diversified classifications, consistent imaging protocols and reliability.

1) IQ-OTH/NCCD

The IQ-OTH/NCCD lung cancer dataset was gathered over a three-month period in the fall of 2019 at the Iraqi Oncology

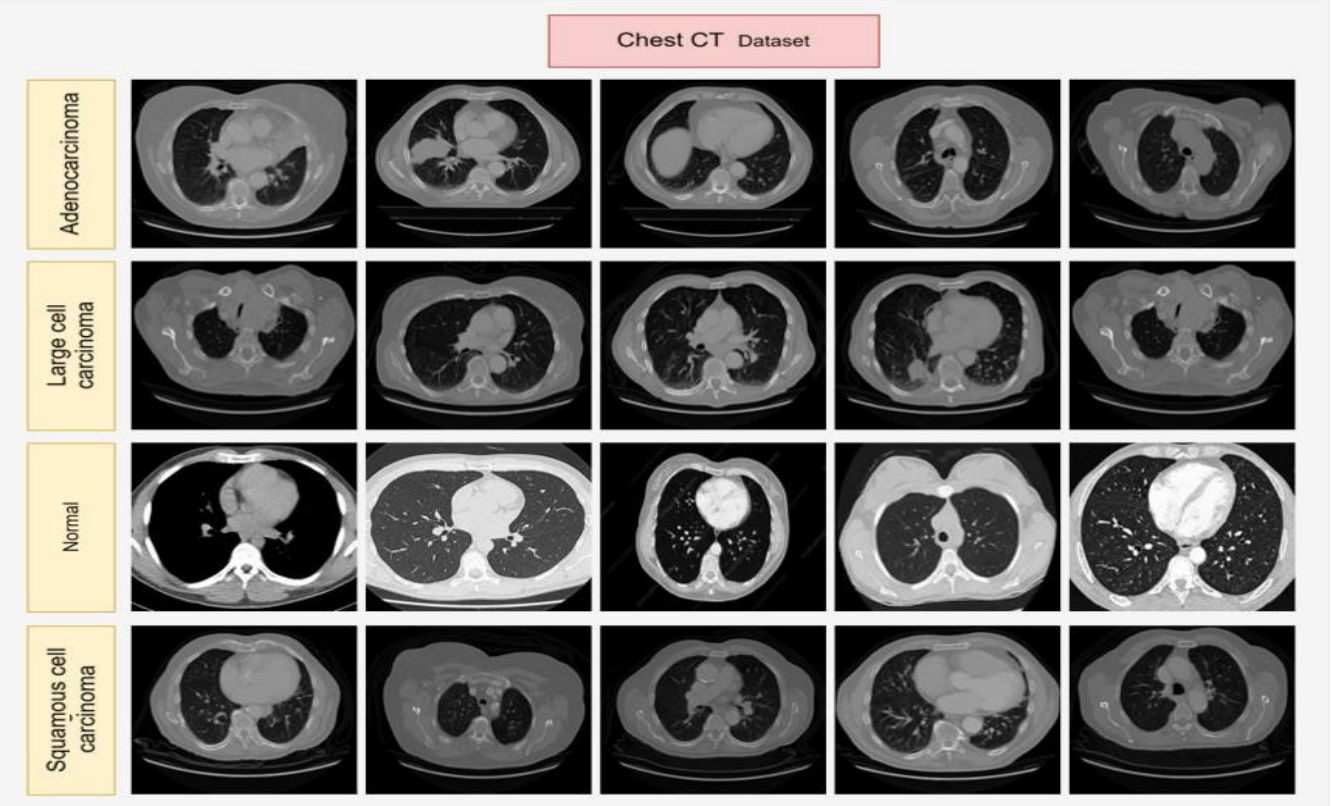

FIGURE 2. Example images for each class from the chest CT dataset.

Teaching Hospital and National Cancer Disease Center. The collection comprises CT images from patients with lung cancer at various stages, as well as healthy people. There are 1197 CT scans from 110 instances. The cases were categorized as 40 malignant, 15 benign, and 55 normal. The pictures were obtained in DICOM format utilizing a Siemens SOMATOM scanner, with a protocol of 120 kV, 1 mm slice thickness, and 350-1200 HU window width. Table 1 shows the distribution of images in the IQ-OTH/NCCD dataset by class [37].

Each scan in the dataset contains between 80 and 200 image slices of the patient's chest taken from different angles. Figure 1 shows sample images from the IQ-OTH/NCCD dataset. Prior to processing, all photos were de-identified, and the study received approval from the individual medical facilities' ethical committees. The cases included in the study were diverse in terms of age, gender, occupation, and region of residence, with the majority of participants coming from the central region of Iraq. The IQ-OTH/NCCD dataset is an important source of data widely used in lung cancer research [37].

2) KAGGLE CHEST CT

The Kaggle Chest CT dataset contains CT images of lung cancers. There are 3 folders in the dataset: training, test, and validation. CT scans are classified into four types: large cell carcinoma, adenocarcinoma, squamous cell carcinoma, and normal. Chest CT contains 1000 CT scans diagnosed with lung cancer. These images are saved as png and jpg. These images are a publicly and freely available dataset with annotations by experienced radiologists.

TABLE 2. Information on the chest CT dataset.

Class	Category	Number of Images
1	Adenocarcinoma	338
2	Large cell carcinoma	187
3	Normal	215
4	Squamous cell carcinoma	260
Total		1000

Table 2 demonstrates the way Chest CT images sed are split into classes. Chest CT provides a large and annotated dataset for training and testing DL models, providing robust and reproducible results for modern medical applications [38]. Example images for each class in the Chest CT dataset are shown in Figure 2.

B. DEEP LEARNING

DL has transformed the area of AI by allowing it to learn from large datasets [39]. One of the most remarkable areas of this

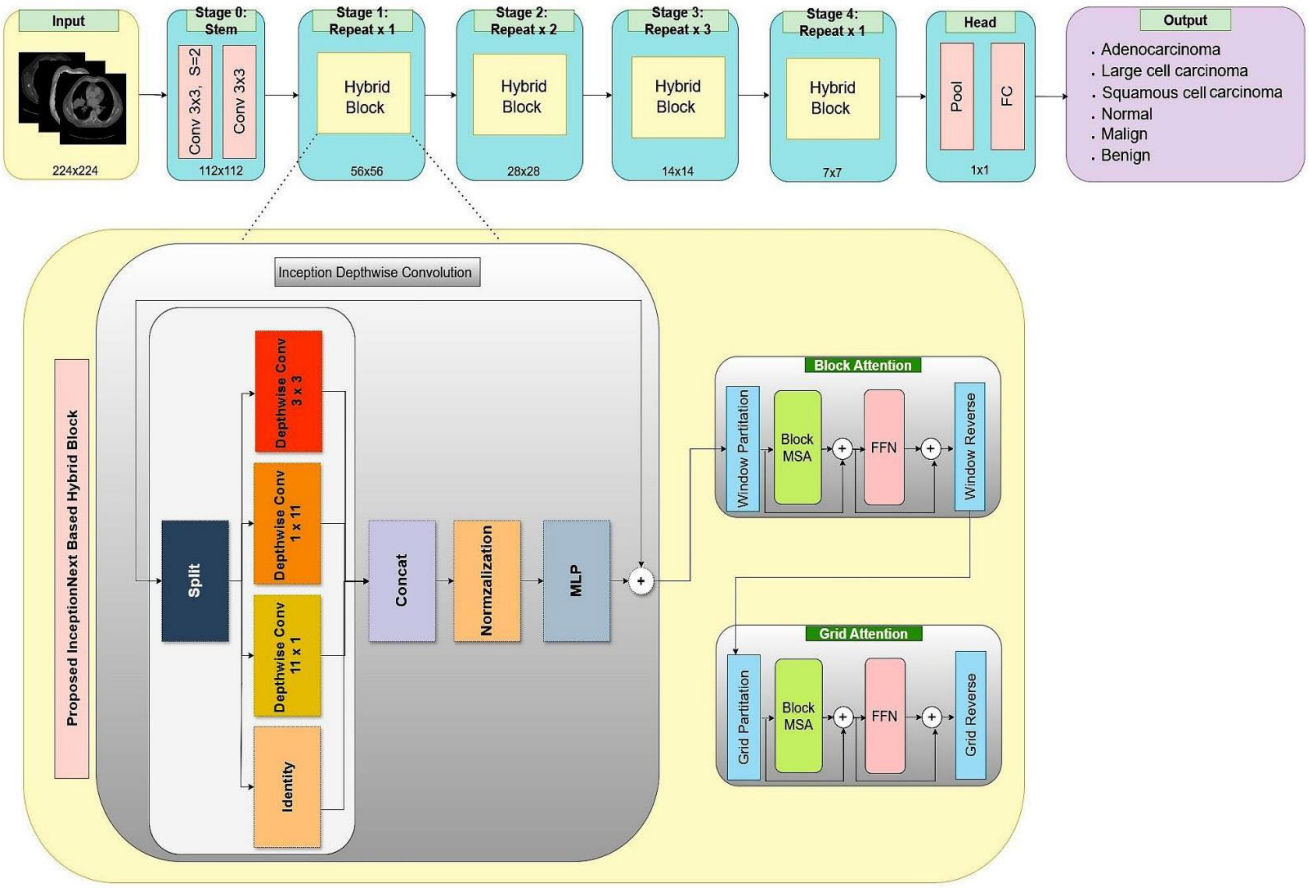

FIGURE 3. Proposed hybrid architecture for lung cancer classification.

transformation is computer vision [40]. Computer vision is used in many different application areas such as face recognition, autonomous vehicles, medical image analysis, and object detection [41]. Among DL architectures, especially CNNs play a highly effective role in analyzing and classifying image data. CNNs allow to identify various features in images with their multi-layered structure. These layers extract important features while reducing the size of the image using convolutional filters and pooling techniques [42]. Due to this hierarchical structure, CNNs learn increasingly abstract representations in images, allowing them to achieve high accuracy rates. However, the inability of CNNs to fully capture the overall information in the image may prevent a full understanding of the contextual relationships between objects.

To overcome these limitations, researchers have developed the ViT model, which is based on the self-attention mechanism and is particularly capable of capturing long-range dependencies [43]. ViT offers a different approach than CNNs by using self-attention and spatial embedding instead of convolutional layers. This model provides a holistic understanding of images with the ability to process both local and global information simultaneously [44]. While CNNs remain the fundamental architecture for computer vision applications, ViTs provide a supplementary approach for more complicated and extensive tasks [45]. By combining these two methods, DL techniques are expected to achieve more impressive results in the field of computer vision in the future.

C. PROPOSED MODEL

Early and precise detection of lung cancer is critical for successful and effective therapy [46]. DL algorithms for detection and classification of malignant tumors have great potential in this field. In general, these algorithms are developed using large datasets [47]. Researchers experiment with small and large model variations to select the most appropriate model for the datasets and problems they are trying to solve. To achieve the best performance, these models must be tuned to the specific dataset. While architectures that are successful on these large datasets can be effective on other datasets, specific optimizations for each dataset are often required [48].

In this study, the Hybrid architecture is restructured for lung cancer diagnosis. The model shown in Figure 3 is derived from the MaxViT architecture to achieve high accuracy rates and provide an efficient and scalable solution. Figure 3 visualizes the detailed architecture and working

principle of the hybrid DL model proposed in this study. The model has been developed to perform complex and sensitive task such as lung cancer detection and presents a hybrid structure by combining both CNN and ViT-based approaches. This architecture aims to exploit both the local feature extraction advantages of convolutional networks and the global information capture capabilities of transformer-based mechanisms.

The architecture is mainly enriched by the integration of InceptionNext and Hybrid blocks. The input layer takes images of size 224 × 224 and first reduces the visual information through a series of convolution operations to extract meaningful features. This process makes it possible to create a more compact feature map with less computational cost. Especially in this layer, called Stage 0 (Stem), successive 3 × 3 convolution operations ensure that local features are preserved.

The model then transitions to Hybrid Blocks, which consist of more than one stage. These blocks combine both convolutional and self-attention mechanisms to extract and process features at different levels. Hybrid blocks were implemented in a total of five stages: Stage 1, Stage 2, Stage 3, Stage 4 and the Head layer. In particular, the Inception Depthwise Convolution section within these blocks performs multi-scale convolution operations, processing various features of the data in parallel. Operations such as Split, Concat, Normalization, and MLP (Multi-Layer Perceptron) support the reconstruction of the data and the creation of a richer representation.

Another key component of the hybrid architecture is Attention Mechanisms. This part is divided into two main components: Block Attention and Grid Attention. The Block Attention mechanism applies multi-head attention (MSA, Multi-Head Self-Attention) at the block level to better understand local features in the image and then enriches this information with FFN (Feed-Forward Neural Network). Grid Attention captures features in the image from a broader perspective and makes sense of grid-level relationships. These mechanisms provide a strong interaction between global and local information.

In the final layer, the features extracted by the Head unit are processed through a pooling and fully connected layer for final classification. The output layer is designed to accurately predict different classes: Adenocarcinoma, squamous cell carcinoma, large cell carcinoma, normal, benign, and malignant.

Scalability is a key feature of DL models, allowing for speed improvements and efficient handling of big datasets. This scalability enhances models' capacity to capture intricate data patterns by increasing the depth, number of nodes, and parameters [49], [50]. Furthermore, correctly scaling the model to meet the given task and dataset allows for the removal of redundant parameters, especially in small datasets, resulting in faster inference and higher efficiency. However, in some cases, overscaling the model can lead to overlearning, which can negatively affect its performance on unseen data. In this study, we provide a novel and improved model for lung cancer diagnosis using the Hybrid architecture. This model provides a more efficient and parameter-efficient solution by scaling different Hybrid structures to the lung cancer dataset. As shown, our goal is to create a lighter model than the Hybrid base architecture with significant adjustments to the number of blocks and channels.

In addition, the self-attention mechanism used in the Hybrid architecture offers an important advantage by allowing neural networks to learn global and local interactions. This can be a practical challenge, especially since the attention mechanism, which is intended to be applied to the entire input space, has a high complexity in terms of computational cost. To address this restriction, a multi-axis attention method known as Max-SA has been created. This technique separates global and local attention components, splits the input feature map into non-overlapping P × P windows, and applies local attention to each window. ''Block attention'' is an approach that decreases the computational cost of computing attention throughout the whole feature space. This technique optimizes model performance by allowing efficient execution of local interactions and supports high accuracy on lung cancer datasets.

1) PROPOSED HYBRID BLOCK

Figure 3 represents the proposed Hybrid block that displays the replacement of the InceptionNeXt block with the InceptionNext block, which is considered one of the most important improvements of the proposed model. The Hybrid block contains a multi-axis attention component consisting of Block Attention and Grid Attention modules. In addition, the original MLP module has been replaced by the GRN-based MLP module, which was first proposed in InceptionNext. The proposed ''Hybrid Block'' is supposed to capture both local and global features by analyzing the input feature maps. The whole process consists of phases like window partitioning, self-attention, and grid partitioning. The InceptionNext layer takes a look at the input feature map and makes use of self-attention in finding the relations between different objects. The output of the Hybrid block is the feature map, which is the result of the processing, for further use in different applications of computer vision.

2) INCEPTIONNEXT BLOCK

The InceptionNeXt architecture is an extended version of the ConvNeXt architecture designed to enhance the DL-based visual tasks performance. The main advantage of this architecture is that it increases efficiency by reducing the high computational cost of large-core deep convolutional processing. InceptionNeXt adopts a structure consisting of small-core convolutions, orthogonal band convolutions, and identity mapping by splitting large-core convolutions into four parallel branches. This approach provides lower memory access costs than traditional large-core processing, while

maintaining model accuracy by providing a large areal sampling space [51], [52].

This proposed architecture is notable for achieving faster training times and higher accuracy rates. These improvements allow DL models to be more efficient and agile, which provides significant advantages when working with large datasets such as lung cancer. The combination of InceptionNeXt and Hybrid for lung cancer detection has the potential to provide more accurate classifications by capturing both local and global features. The InceptionNeXt block is shown in Figure 4.

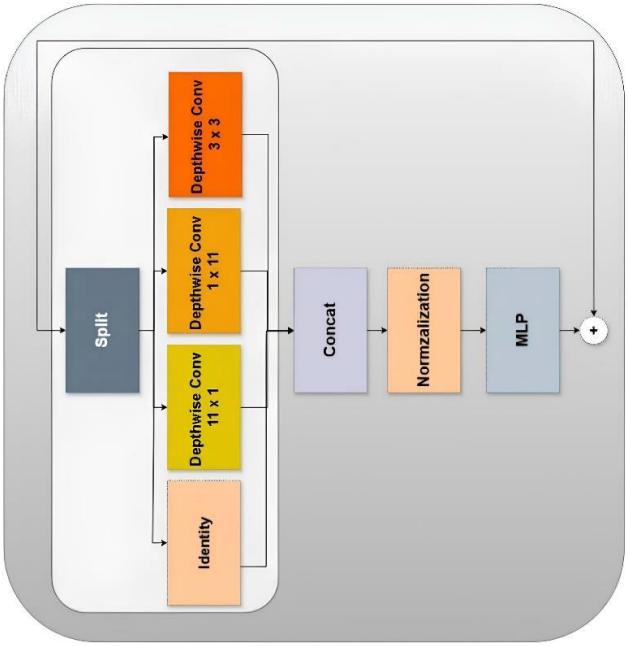

FIGURE 4. InceptionNeXt block.

3) MULTI-AXIS ATTENTION

The Hybrid model represents one of the most advanced approaches in transformer-based image processing, demonstrating exceptional performance. A key feature of this design is the self-attention mechanism, which surpasses traditional local convolution techniques by enabling the network to capture both global and contextual relationships within the data. One notable variant utilized in this architecture is "relative attention," which is specifically designed to model the relative positions and interactions among elements in a sequence with greater precision. This capability is particularly advantageous in complex tasks where spatial dependencies play a critical role.

However, the quadratic computational complexity of standard self-attention poses a significant challenge when applied to large-scale inputs, as it can quickly become computationally prohibitive. To overcome this limitation, the Hybrid model incorporates a multi-axis attention mechanism referred to as Max-SA. This innovative approach effectively balances computational efficiency and accuracy by decomposing the attention process into separate components that focus on global and local features independently. By partitioning the input into manageable subspaces, Max-SA ensures that the model maintains its capacity to capture broad contextual relationships while efficiently processing fine-grained details, making it highly effective for complex image analysis tasks.

$$\begin{aligned} &Relative\ Attention\ (Q, K, V) \\ &= softmax\left(\frac{QK^T}{\sqrt{d}} + B\right) V \end{aligned} \tag{1}$$

$$\begin{aligned} &Block : (H, W, C) \\ &\rightarrow \left(\frac{H}{P}xP, \frac{W}{P}xP, C\right) \rightarrow \left(\frac{HW}{P^2}, P^2, C\right) \end{aligned} \tag{2}$$

$$\begin{aligned} &Grid : (H, W, C) \\ &\rightarrow \left(Gx\frac{H}{G}, Gx\frac{W}{P}, C\right) \rightarrow \left(G^2, \frac{HW}{G^2}, C\right) \\ &\rightarrow \left(\frac{HW}{G^2}, G^2, C\right) \end{aligned} \tag{3}$$

The Max-SA method divides the feature map into non-overlapping $P \times P$ windows, enabling the application of a self-attention mechanism to the local spatial dimensions within each window. This strategy is specifically designed to reduce the computational complexity associated with applying attention across the entire feature space. Known as "block attention," this approach enhances the model's ability to capture and process local interactions effectively. As detailed in equations (1) to (3), block attention operates on an input $x \epsilon R^{HxWxC}$ to focus on extracting localized features. Complementing this, the grid attention module, described in equation (5), addresses global feature exploration by modeling broader contextual relationships within the feature map. Together, these mechanisms ensure a balance between efficiency and accuracy in feature extraction.

$$\begin{aligned} &x \leftarrow x + Unblock\ (RelAttention\ (Block\ (LN\ (x)))) \ and\ x \\ &\leftarrow x + GRN\ (LN\ (x)) \end{aligned} \tag{4}$$

$$\begin{aligned} &x \leftarrow x + Ungrid\ (RelAttention\ (Grid\ (LN\ (x)))) \ and\ x \\ &\leftarrow x + GRN\ (LN\ (x)) \end{aligned} \tag{5}$$

The Q, K, and V input formats employed in the Relative Attention mechanism are omitted in this discussion for brevity. Key components of this process include Layer Normalization (LN) and GRN, an innovative variation of the Multi-Layer Perceptron (MLP). Grid Attention leverages a fixed, uniform GxG grid structure to enable sparse global attention, partitioning tensors into windows of varying sizes. As described in Equation (4), this approach facilitates global spatial mixing of tokens by extending the GxG self-attention grid, enabling efficient processing of global features.

The proposed Max-SA module offers an efficient alternative to existing attention mechanisms, maintaining the same number of parameters and FLOPs while simplifying implementation. Unlike traditional designs, it does not rely on masking, padding, or cyclic shifting, which streamlines its

integration into models. In contrast to block attention, which is restricted to localized windows, grid attention employs a sparse, uniform grid structure to emphasize pixel interactions across the entire 2D scene. Both attention methods utilize fixed attention principles, facilitating spatial interactions with similar shades and achieving linear complexity proportional to input size, thereby enhancing computational efficiency without compromising performance. When determining the configuration of the hybrid blocks, block attention and grid attention mechanisms are used together to ensure a balanced extraction of local and global features. Block attention works on small $P \times P$ windows to capture local features in more detail, while grid attention uses a fixed GxG grid structure to model contextual relationships over larger areas. This combination minimizes computational cost and improves accuracy.

IV. RESULT AND DISCUSSION

A. EXPERIMENTAL DESIGN

In this study, the experiments were conducted on a computer running the Ubuntu 22.04 operating system. DL models were trained and tested on a high-performance system equipped with an Intel 13,600 K processor, an NVIDIA RTX 4090 GPU, and 32 GB of DDR5 RAM. To ensure consistency, the most recent stable version of the PyTorch framework with NVIDIA CUDA support was utilized, and all models were evaluated under identical computational settings and parameters.

B. PERFORMANCE METRICS

In DL, metrics for performance are critical for determining algorithm efficacy. These indicators give useful insights for optimizing and improving model performance, identifying potential mistakes and biases, and ensuring correct predictions. However, focusing exclusively on accuracy may not necessarily result in a full review, especially when dealing with skewed datasets. Metrics like accuracy, recall, and the F1 score provide a more in-depth study by concentrating on the model's ability to produce correct positive predictions and identify all relevant positive occurrences. These metrics are especially important in applications where class distributions are skewed, as high accuracy might mask poor performance on minority classes [54].

Accuracy quantifies the proportion of correct predictions, including both positives and negatives, relative to the total number of predictions. While it provides a general overview, accuracy may not adequately reflect the model's performance on specific classes, particularly in imbalanced datasets. Precision focuses on the ratio of true positive predictions to all predicted positives, highlighting the model's capability to minimize false positives. Recall, also known as sensitivity, measures the proportion of true positives correctly identified among all actual positive cases, emphasizing the model's ability to detect relevant instances. The F1 score combines precision and recall as their harmonic mean, offering a balanced evaluation of performance. This metric is especially valuable in scenarios where both precision and recall are critical, and data imbalance is a concern.

Mathematically, these metrics are expressed as follows: Accuracy refers to the proportion of correct predictions out of the total number of predictions made. Precision is applied to find the ratio in regards to correctly predicted positive observations against all those that could be predicted as positive ones. Recall, also scientifically known as sensitivity, relies heavily on the ratio of total positive instances identified correctly during the experiment to the amount of actual positive instances taken together. Lastly, the F1 score represents the harmonic mean of precision and recall, providing a balanced measure of a model's performance. These equations, detailed in (6)-(9), allow for a comprehensive assessment of the model's performance. Together, these metrics enable a nuanced evaluation, addressing both prediction quality and the challenges posed by class imbalances effectively [55].

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Number\ of\ Total\ Prediction} \tag{6}$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{7}$$

$$Recall = Sensitivity = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{8}$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \tag{9}$$

C. DATA PREPROCESSING AND DATA AUGMENTATION

Data preparation is a vital step in improving DL models' performance and generalization capabilities. The phases in this procedure involve data partitioning into training, validation, and test sets, as well as normalization, noise reduction, and outlier handling. Instead of the two-cluster partitioning usually utilized in the literature, this study opts for a three-cluster structure: training, validation, and testing. This strategy is necessary to more accurately evaluate the model's performance and to avoid overlearning. The training set was utilized to optimize the model's parameters and accounted for 70% of the dataset. The validation set (10%) was used to evaluate its generalization ability and to optimize the hyperparameters, while the test set (20%) measured the model's performance on data it had not seen in training. This threefold separation allows for a more objective and accurate evaluation of the model's performance. Since the Chest CT dataset contains predefined training, validation and test subgroups, no changes were made to these subgroups to ensure a fair and standardized comparison and the dataset was used as is. Tables 3 and 4 show the class distribution for the Chest CT and IQ-OTH/NCCD datasets. Data augmentation methods were applied to the Chest CT and IQ-OTH/NCCD lung cancer detection datasets. Data augmentation is an important method to increase and consolidate the generalization ability of models working with limited data. Especially in medical imaging, where data quality varies, these methods are of great importance. In this study, data augmentation techniques such

TABLE 3. Class-wise distribution of the chest CT dataset.

Class	Category	Number of Images		
		Test	Train	Validation
1	Adenocarcinoma	120	195	23
2	Large cell carcinoma	51	115	21
3	Normal	54	148	13
4	Squamous cell carcinomave	90	155	15
Total		315	613	72

TABLE 4. Class-wise distribution of the IQ-OTH/NCCD dataset.

Class	Category	Number of Images		
		Test	Train	Validation
1	Malign	112	392	56
2	Benign	24	84	12
3	Normal	83	291	41
Total		219	767	109

as cropping, rotation, translation, scaling, and random noise addition were used. These techniques aim to increase the diversity of the dataset and reduce overlearning. Especially for models working with limited data, data augmentation significantly improved the ability of the model to generalize to unseen data samples.

D. TRANSFER LEARNING

Transfer learning is a highly effective technique for enhancing the performance and efficiency of DL models. By utilizing the weights and representations of a model previously trained on a related task, this method enables the reuse of acquired knowledge in a new domain. This approach is particularly beneficial in scenarios where transferring features from models trained on extensive datasets to tasks with limited data enables efficient problem-solving with reduced data and resource requirements. It shortens training time, improves model accuracy, and enhances generalization capabilities, making it widely applicable across various domains.

In particular, transfer learning has proven to be invaluable in domains such as medical image analysis and classification, where datasets are often limited and imbalanced. For instance, datasets used for cancer detection are frequently small and lack balance across classes. Utilizing weights from models trained on large-scale datasets facilitates the effective learning of task-specific features. In this study, we utilized pre-trained weights from a model initially trained on the ImageNet dataset, comprising millions of images and diverse classes. These weights were fine-tuned to adapt to the smaller Chest CT and IQ-OTH/NCCD datasets. This method takes advantage of the pre-trained model's robust generalization and feature extraction capabilities, leading to faster training and reduced computational overhead. Moreover, this technique improved the model's performance on limited datasets, facilitating higher accuracy while requiring fewer computational resources.

E. TRAINING DETAILS

Effective deep learning model training depends on the strategic combination of methods and parameter fine-tuning to improve accuracy and computational efficiency. Techniques like data augmentation and transfer learning are particularly crucial when dealing with datasets that are either limited in size or lack diversity. Transfer learning utilizes pre-trained models, such as those developed on ImageNet, for new tasks, enhancing generalization capabilities. Meanwhile, data augmentation enriches dataset variability by introducing transformations like rotation, scaling, and noise addition, which bolster model robustness and reduce the risk of overfitting.

Key factors, including image resolution, batch size, epoch count, optimizer selection, and learning rate, were meticulously optimized to strike an optimal balance between performance and efficiency. To address overfitting, a weight decay parameter was employed, while adaptive learning rate schedules facilitated faster convergence with improved stability. Early training phases incorporated gradual learning rate increases to stabilize the optimization process, with the momentum parameter ensuring consistent progression. Additional measures, such as warm-up epochs and initial learning rate adjustments, were implemented to mitigate abrupt parameter changes during the initial stages.

To ensure consistency and reproducibility, all models underwent training for 300 epochs under standardized conditions. These included an input resolution of 224×224 pixels, an initial learning rate of 0.01, a base learning rate of 0.1, a momentum value of 0.9, weight decay set to 2.0e-05, and the use of the SGD optimizer. Five warm-up epochs with a starting learning rate of 1.0e-05 were incorporated into the training process. Additionally, consistent data augmentation methods, such as scaling, rotation, and inversion, were applied uniformly across all models.

This comprehensive experimental design, which integrated ImageNet pre-trained weights during the transfer learning phase, provided a robust foundation for systematic and fair model evaluation. By utilizing advanced computational resources and optimized hyperparameters, this approach effectively trained and assessed the Proposed Model, addressing the complexities of lung cancer detection and delivering outstanding performance across diverse datasets.

F. EXPERIMENTAL RESULTS

In this study, the lung cancer detection performance of the proposed model is evaluated on two different datasets, IQ-OTH/NCCD and Chest CT, and the results obtained are compared. The performance of the model is analyzed in terms of performance criteria such as accuracy, sensitivity, precision and F1 score. In these criteria, it is shown that it performs superior to the existing methods in the literature. The success of the model is characterized by its consistent

TABLE 5. Results of CNN-based and ViT-based models on IQ-OTH/NCCD dataset.

Model	Accuracy	Precision	Recall	F1-Score
ResNet50	0.9315	08851	0.8629	0.8729
DenseNet169	0.9406	0.9349	0.8491	0.8780
EfficientNetv2-Medium	0.9635	0.9357	0.9185	0.9266
ConvNeXt-Base	0.9772	0.9811	0.9306	0.9515
InceptionNeXt-Base	0.9772	0.9694	0.9404	0.9535
MobileViT-Small	0.9726	0.9540	0.9364	0.9446
ConViT-Base	0.9635	0.9437	0.9086	0.9237
Swin-Base	0.9772	0.9425	0.9799	0.9582
MaxViT-Base	0.9726	0.9650	0.9375	0.9497
DeiT3-Base	0.9817	0.9642	0.9642	0.9642
Proposed Model	**0.9954**	**0.9967**	**0.9960**	**0.9912**

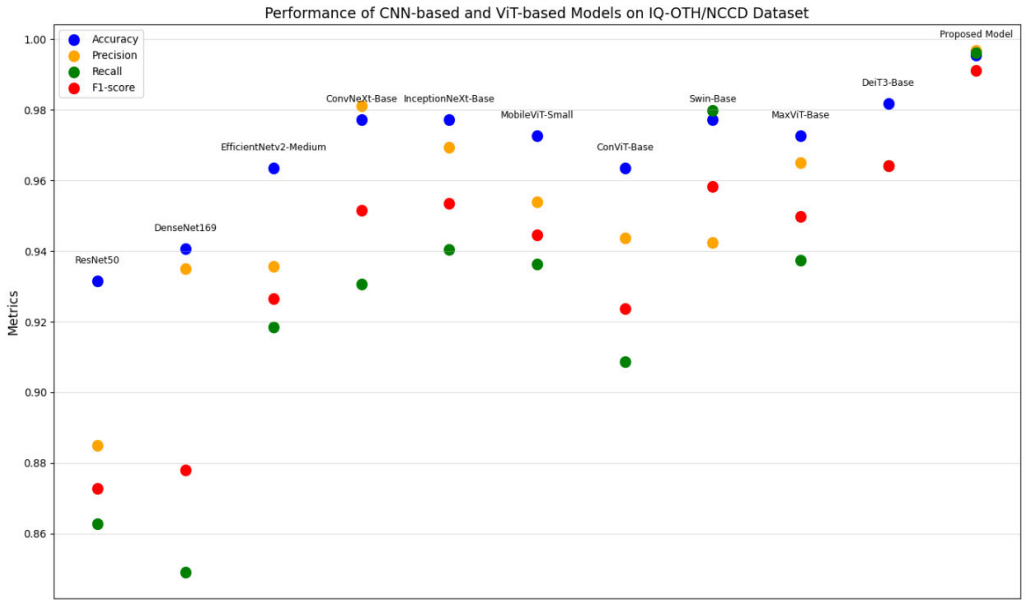

FIGURE 5. Evaluation of the proposed model for IQ-OTH/NCCD dataset against CNN and ViT.

results in two different data sets, its general validity and its minimal dependency on specific data sets. In addition, all models compared in the study were trained on the same data sets and under the same experimental conditions. This approach has been applied with special care to ensure the consistency and reliability of the performance evaluations. The main reason for choosing the proposed hybrid approach is to provide a more comprehensive and effective analysis for lung cancer diagnosis by combining the local feature extraction capability of CNNs with the global contextual information capture capability of ViTs. The lightweight architecture of the model allowed not only computational efficiency but also high accuracy rates to be achieved. In particular, the same experimental conditions were used to realize the study in a fair environment. This allows for an objective comparison of the results and demonstrates the truly superior performance of the model.

1) RESULTS FOR IQ-OTH/NCCD DATASET

The model used in this study is characterized by its lightweight structure, which plays a key role in achieving high performance. Due to its compact design, the proposed

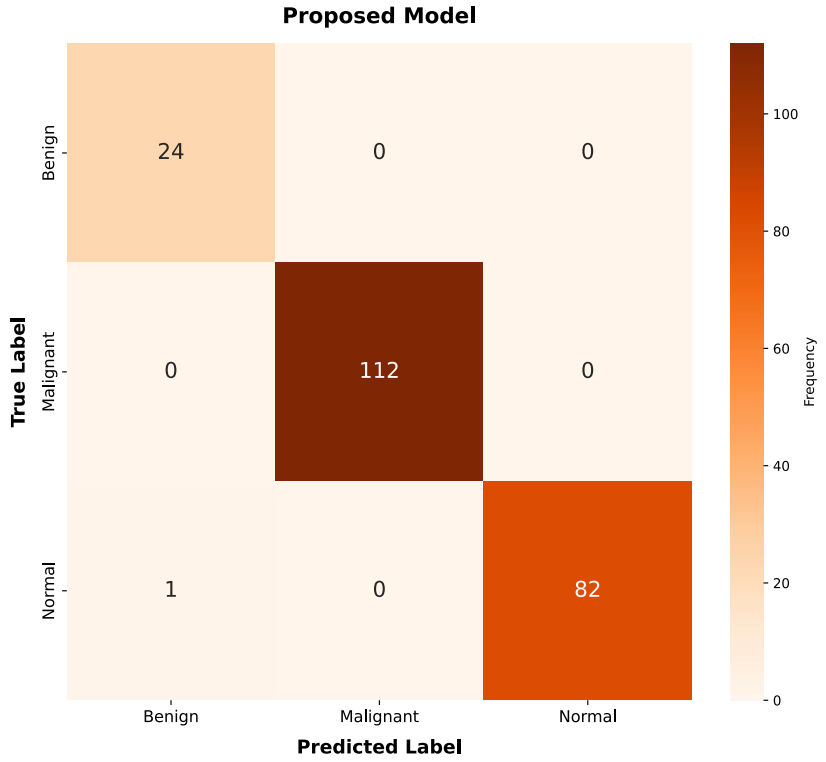

FIGURE 6. Confusion matrix for IQ-OTH/NCCD dataset.

model not only achieves computational efficiency but also performs well in the challenging task of lung cancer detection. Reducing the memory usage by optimizing the number of parameters and at the same time increasing the accuracy clearly demonstrates the strengths of the proposed approach and its success in practice.

Table 5 shows the results of CNN-based and ViT-based models on the IQ-OTH/NCCD dataset. This table compares the performance of several models in terms of accuracy, precision, recall, and f1-score. Our proposed model stands out from the other models due to its lightweight construction and high success rate. Table 5 shows the performance of several CNN and ViT models on the IQ-OTH/NCCD dataset. The suggested model exceeds the previous models in all criteria, with 99.54% accuracy, 99.67% precision, 99.60% recall, and a 99.12% F1 score. These results show that the proposed model not only provides high efficiency due to its lightweight structure but also demonstrates superior performance with high accuracy rates. Traditional CNN-based models, such as ResNet50 and DenseNet169, showed inferior performance with accuracy rates of 93.15% and 94.06%, respectively. These models are particularly weak in sensitivity and F1 score-based results. For example, the ResNet50 model was limited to 86.29% accuracy, suggesting that it is particularly inadequate for small or complex datasets. Although DenseNet169 provides better precision (93.49%) than the other CNN models, its low precision (84.91%) highlights its limitations. EfficientNetv2-Medium was more competitive among the CNN-based models, achieving 96.35% accuracy, 93.57% precision and 91.85% sensitivity. However, it lagged behind ViT-based models. ViT-based models, especially Swin-Base, DeiT3-Base, and InceptionNeXt-Base, stood out for their high accuracy rates and balanced metric values. Swin-Base was one of the prominent models with 97.72% accuracy and 97.99% sensitivity, and it performed better than other ViT-based models, especially in terms of sensitivity. However, it lagged behind the proposed model in F1 score with 95.82%. DeiT3-Base, on the other hand, showed a very balanced performance, achieving 98.17% accuracy and 96.42% precision, sensitivity and F1-score. One of the most important factors for the success of the proposed model is its lightweight and optimized structure. The model provides high accuracy and precision while reducing the number of parameters and computational cost, which is a great advantage especially for systems with limited resources. In addition, the high sensitivity rate of this model (99.60%) shows that false negatives are minimal, which makes a significant contribution to accurate diagnosis in a critical area such as lung cancer. In conclusion, the results in Table 5 clearly demonstrate that the proposed model outperforms both CNN and ViT-based models. In particular, the high values obtained in accuracy, precision, sensitivity and F1 score show that the proposed model provides a superior solution not only in theory but also in practical applications.

Figure 5 shows a dot plot of the models' performances on the IQ-OTH/NCCD dataset in terms of accuracy, precision,

TABLE 6. Results of CNN-based and ViT-based models on chest CT dataset.

Model	Accuracy	Precision	Recall	F1-Score
ResNet50	0.8190	0.8291	0.8535	0.8271
DenseNet169	0.9111	0.9245	0.9266	0.9219
EfficientNetv2-Medium	0.8190	0.8334	0.8631	0.9341
ConvNeXt-Base	0.9238	0.9205	0.9356	0.9270
InceptionNeXt-Base	0.9397	0.9367	0.9488	0.9413
MobileViT-Small	0.9524	0.9518	0.9600	0.9539
ConViT-Base	0.9524	0.9484	0.9613	0.9526
Swin-Base	0.9683	0.9707	0.9710	0.9707
MaxViT-Base	0.9460	0.9498	0.9586	0.9537
DeiT3-Base	0.9524	0.9467	0.9627	0.9526
Proposed Model	**0.9841**	**0.9861**	**0.9835**	**0.9848**

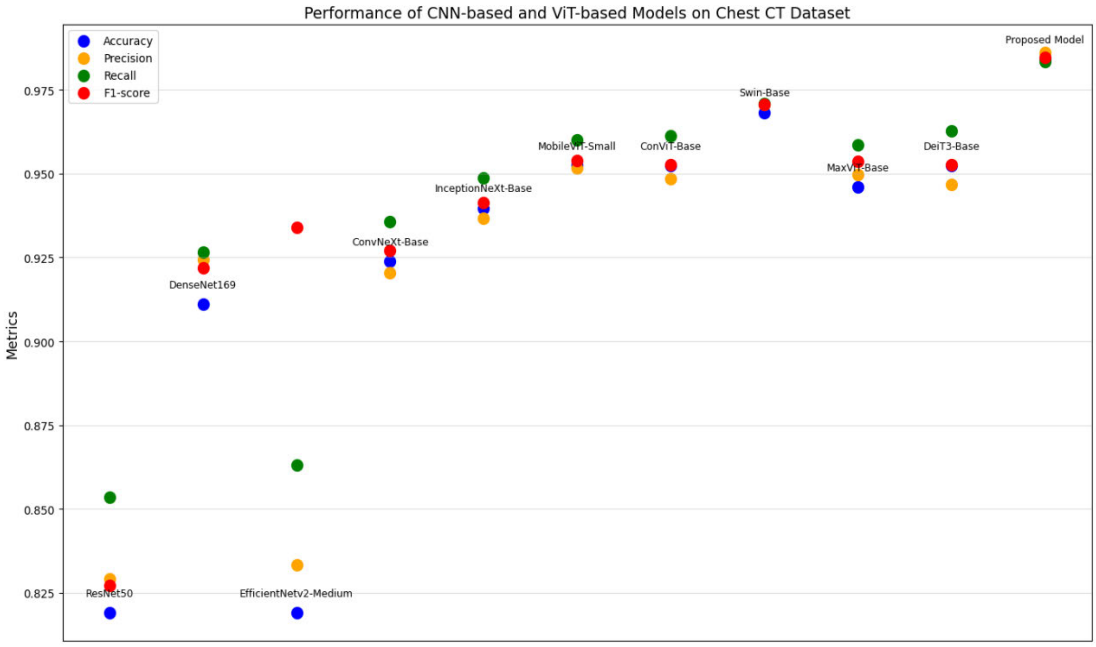

FIGURE 7. Evaluation of the proposed model for Chest-CT dataset against CNN and ViT.

recall, and f1-score. This graph shows that the suggested model beats the existing CNN and ViT-based models across all measures.

Figure 6 depicts the confusion matrix, which measures class-wise performance on the IQ-OTH/NCCD dataset. This matrix illustrates the number of correct and erroneous classifications for each class in detail, demonstrating the suggested model's capacity to classify malignant, benign, and normal classes accurately.

In the benign class, all 24 samples were correctly classified and the model achieved 100% accuracy in this class. In the Malignant class, all 112 samples were correctly identified, demonstrating the excellent success of the model in detecting critically important malignant tumors. In the normal class, 82 out of 83 samples were correctly classified, with only one sample incorrectly classified as benign. This affected the overall success of the model by a very small percentage. These results show that the proposed model has a balanced

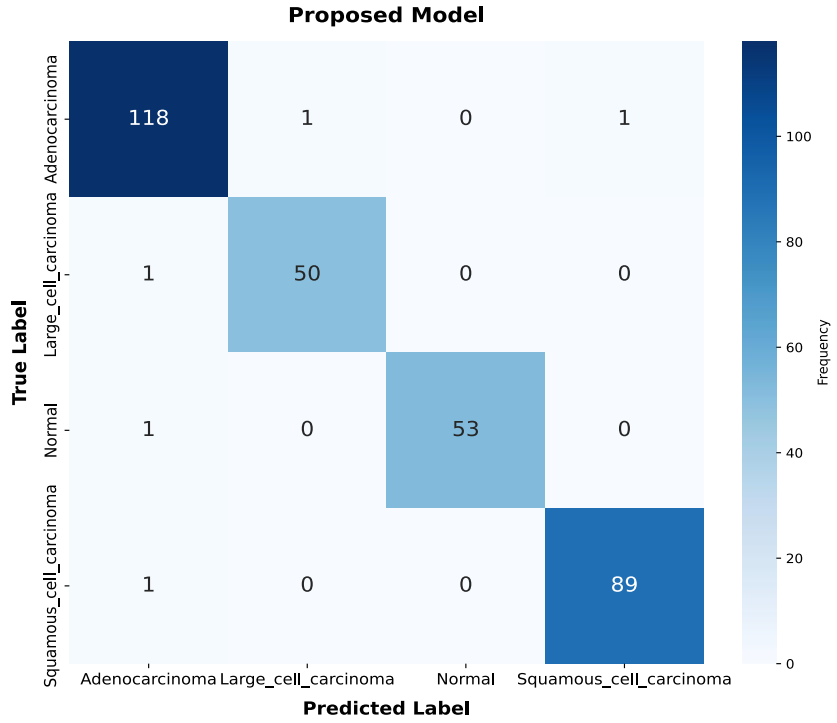

FIGURE 8. Confusion matrix for Chest CT dataset.

and high performance in all three classes. Especially the 100% accuracy in malignant tumors increases the usability of the model in clinical applications. The lightweight nature of the model provides both computational efficiency and strong discrimination ability between classes, with only one misclassification.

2) RESULTS FOR CHEST CT DATASET

The model proposed in this paper, with its lightweight structure, not only improves computational efficiency but also achieves very high performance. The compact and optimized design of the model made it possible to achieve superior performance with less computational resources.

Table 6 compares the outcomes of several models trained on the chest CT dataset using important performance parameters like as accuracy, precision, recall, and the F1 score. This table compares the performance of the proposed model to other CNN and ViT-based models. According to the obtained results, the proposed model shows significantly better performance than other models. In particular, it yields the highest values in terms of accuracy (98.41%), precision (98.61%), sensitivity (98.35%), and F1 score (98.48%), demonstrating its superior performance.

Analyzing the performance of the other models in Table 6, one can observe that the Swin-Base model gives the closest results to the proposed model in terms of accuracy (96.83%), precision (97.07%), sensitivity (97.10%), and F1 score (97.07%). However, even the Swin-Base model could not match the metrics provided by the proposed model. Similarly, the InceptionNeXt-Base and MobileViT-Small models also showed remarkable performance but lagged behind the proposed model, especially in terms of sensitivity and F1 score. Among the CNN-based models, DenseNet169 outperformed the other CNN models with an accuracy of 91.11%. However, this model still has a very low performance compared to the proposed model. On the other hand, models such as ResNet50 and EfficientNetv2-Medium had an accuracy of 81.90% and similarly poor performance in other metrics. This shows that classical CNN-based approaches have a more limited impact against ViT-based models and the proposed model. One of the most important factors behind the success of the proposed model is that it offers a more efficient learning process due to its lightweight structure. By optimizing the memory and processing power requirements, the model reduces the computational cost and achieves high accuracy. In particular, the high recall value shows that the proposed model minimizes the false negative rate and thus provides a critical advantage in early diagnosis of the disease. In addition, the high precision of the model emphasizes that the false positive rate is low and thus unnecessary treatments can be avoided. The consistently high F1 score demonstrates the overall success of the model.

Figure 7 depicts the models' performance on the Chest CT dataset in terms of accuracy, precision, recall, and F1 score metrics using a dot plot. This graph allows you to easily compare the performance of different models and clearly illustrates that the suggested model outperforms the other models.

Figure 8 depicts the confusion matrix, which analyzes the class-wise performance of the proposed model on the Chest CT dataset. The image depicts the right and wrong classifications for the four types (large cell carcinoma, adenocarcinoma, normal and squamous cell cancer). The model's ability to differentiate between classes is assessed by comparing the correct predictions and error rates for each class.

In the Adenocarcinoma class, 118 cases were correctly classified and only 2 cases were incorrectly predicted. One of these incorrect predictions was classified as Large_cell_carcinoma and the other as Squamous_cell_carcinoma. In the large_cell_carcinoma class, 50 samples were correctly predicted and only one sample was misclassified as adenocarcinoma. These results show that the model achieved high accuracy in both classes.

In the Normal class, all cases were correctly predicted (100% success), showing that the model can perfectly discriminate healthy individuals. In the Squamous_cell_carcinoma class, 89 samples were correctly classified and only one sample was incorrectly predicted as adenocarcinoma. The accuracy rate in this class is also quite high.

Looking at the overall results in Figure 8, it can be seen that the proposed model has a balanced and successful performance in all classes. The model stands out for its low misclassification rates and especially for its 100% accuracy rate in the Normal class. The low error rates observed in the Adenocarcinoma, Large_cell_carcinoma and Squamous_cell_carcinoma classes indicate that the model can effectively discriminate different types of lung cancer. These results demonstrate that the proposed model exhibits high accuracy and consistency on the chest CT dataset, providing a potentially useful solution for clinical applications.

3) INTERPRETABILITY OF THE PROPOSED MODEL THROUGH GRAD-CAM ANALYSIS

Gradient-weighted Class Activation Mapping (Grad-CAM) is a vital interpretability tool in DL, allowing for the visualization of regions within an input image that contribute most significantly to the model's predictions. By computing the gradients of a target class relative to the final convolutional layer, Grad-CAM produces heatmaps that provide a clear depiction of the model's focal areas. This method is particularly valuable in medical imaging tasks, such as lung cancer detection, as it ensures that the model's decision-making process aligns with clinically significant features, fostering transparency and trustworthiness [56].

Figure 9 showcases Grad-CAM heatmaps for three representative CT scan cases: benign, malignant, and normal. The upper row contains the original ground truth images, while the lower row presents the corresponding Grad-CAM visualizations. These heatmaps illustrate the regions the model deemed most relevant for its classification decisions, offering a deeper understanding of its behavior.

In the benign case (first column), the Grad-CAM visualization highlights specific localized regions within the lung, corresponding to minor irregularities or benign nodules. This focused activation indicates the model's ability to identify subtle but clinically relevant features, demonstrating a nuanced understanding of benign conditions that aligns with expert radiological assessment.

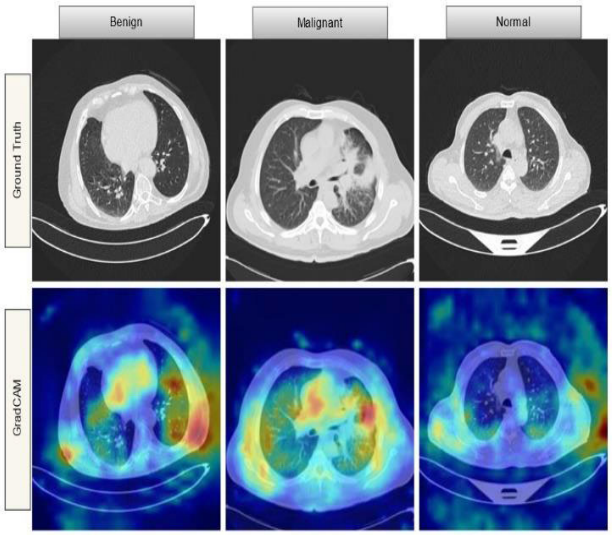

FIGURE 9. Grad-CAM heat maps for CT scan case.

The malignant case (second column) reveals a more extensive activation in the Grad-CAM heatmap, covering large and abnormal regions within the lung tissue. These highlighted areas correspond to key clinical indicators of malignancy, such as prominent nodules, masses, or diffuse patterns. The intensity and breadth of the activation confirm the model's capability to detect critical features with high confidence, underscoring its effectiveness in identifying malignancies.

For the normal case (third column), the Grad-CAM heatmap displays minimal activation, consistent with the absence of abnormalities. The lack of significant highlighted regions reflects the model's precision in distinguishing normal cases from those with pathology. This indicates that the model effectively ignores irrelevant features, reinforcing its reliability in classifying normal lung scans accurately.

The Grad-CAM analysis provides essential insights into the model's interpretability, confirming that it focuses on clinically meaningful areas for its predictions. This transparency not only validates the model's predictions but also enhances its credibility as a reliable tool for lung cancer detection. By highlighting critical regions that align with established medical knowledge, the analysis establishes the model's potential for real-world clinical applications.

4) EVALUATION OF MODEL PERFORMANCE (BASELINE VS PROPOSED MODELS)

To validate the effectiveness of the proposed model and establish its superiority over state-of-the-art (SOTA) models,

TABLE 7. Paired t-test results for statistical significance.

Analysis	IQ-OTH/NCCD Dataset		Chest CT Dataset	
Model	Accuracy p-value	F1-Score p-value	Accuracy p-value	F1-Score p-value
ResNet50	0,0031	0,0052	0,0024	0,0047
DenseNet169	0,0028	0,0046	0,0021	0,0039
EfficientNetv2-Medium	0,0015	0,0033	0,0011	0,0027
ConvNeXt-Base	0,0012	0,0025	0,001	0,0023
InceptionNeXt-Base	0,001	0,0022	0,0009	0,0019
MobileViT-Small	0,0009	0,0018	0,0008	0,0017
Swin-Base	0,0008	0,0016	0,0007	0,0014
DeiT3-Base	0,0006	0,0013	0,0005	0,0012
Proposed Model	< 0.0001	< 0.0001	< 0.0001	< 0.0001

we performed paired t-tests for each model across two datasets: IQ-OTH/NCCD and Chest CT. The metrics under evaluation included Accuracy and F1-Score. This rigorous statistical analysis aimed to identify whether the performance improvements of the proposed model are statistically significant or could be attributed to random variation.

Table 7 provides a detailed summary of the p-values derived from the paired t-tests for Accuracy and F1-Score across both datasets. A p-value below 0.05 was considered statistically significant, highlighting meaningful performance differences between the proposed model and baseline models.

The paired t-test results in Table 7 provide a comprehensive evaluation of the statistical significance of the proposed model's performance improvements compared to baseline models. These findings underscore the robustness and reliability of the proposed model across two datasets, IQ-OTH/NCCD and Chest CT, with distinct characteristics and challenges.

The proposed model exhibits statistically significant superiority over all baseline models, including ResNet50, DenseNet169, and ConvNeXt-Base. The p-values for both Accuracy and F1-Score comparisons are consistently below 0.01, confirming that the observed differences are not attributable to random variation. While some baseline models, such as Swin-Base and DeiT3-Base, demonstrate competitive results, their performance still falls short of the proposed model's metrics. This demonstrates the effectiveness of the proposed model in handling the complex features of the IQ-OTH/NCCD dataset, characterized by diverse image patterns.

The analysis on the Chest CT dataset reveals similar trends, with the proposed model significantly outperforming all baseline models. The p-values for Accuracy and F1-Score comparisons remain below 0.01, further reinforcing the statistical significance of its performance gains. This consistent performance across datasets highlights the model's adaptability to various data distributions and its capability to maintain high levels of precision and reliability in medical imaging tasks.

TABLE 8. Performance comparison of model variants with grid and block attention.

Model	Accuracy of IQ-OTH/NCCD dataset	Accuracy of Chest CT dataset
Baseline Model	0.9772	0.9397
Baseline Model + Grid Attention	0.9819	0.9494
Baseline Model + Bock Attention	0.9820	0.9525
Proposed (Baseline + Grid + Block)	0.9954	0.9841

Across both datasets, the statistical significance of the proposed model's performance highlights its superior capability in feature extraction and classification tasks. The consistently low p-values demonstrate that the performance differences are meaningful and not coincidental. Moreover, the results validate the efficacy of the model's design, which integrates CNN and ViT architectures to effectively capture both local and global features.

These findings suggest that the proposed model holds substantial potential for broader applications in real-world scenarios, particularly in domains where high accuracy and reliability are critical, such as medical imaging and diagnostics. The robust statistical validation provides a solid foundation for further research and development of hybrid architectures that utilize the strengths of CNNs and ViTs.

5) ABLATION STUDY: COMPONENT-WISE EVALUATION OF THE PROPOSED METHOD

To comprehensively assess the contributions of key architectural components in the proposed model, we performed a detailed component-wise evaluation. This analysis aims to isolate the effects of Grid Attention and Block Attention mechanisms, both individually and in combination, to determine their respective roles in improving the model's performance. By analyzing their impact on two datasets—IQ-OTH/NCCD and Chest CT—we validate whether the integration of these mechanisms effectively enhances the model's ability to capture fine-grained local details and global contextual patterns.

The evaluation included four configurations: the baseline model, the baseline model with Grid Attention, the baseline model with Block Attention, and the full proposed model, which integrates both Grid and Block Attention mechanisms. The results are summarized in Table 8, showcasing the incremental benefits provided by these mechanisms and their synergy in the proposed architecture.

As seen in Table 7, the baseline model demonstrates a strong foundational performance with accuracies of 0.9772 and 0.9397 on the IQ-OTH/NCCD and Chest CT datasets, respectively. However, when Grid Attention is

incorporated, there is a notable increase in accuracy on both datasets. This mechanism enhances the model's ability to capture global patterns by focusing on large-scale contextual features, resulting in a 0.0047 improvement on IQ-OTH/NCCD and a 0.0097 improvement on Chest CT.

Similarly, adding Block Attention to the baseline model leads to further performance gains. This mechanism focuses on local, fine-grained details, boosting accuracies to 0.9820 on IQ-OTH/NCCD and 0.9525 on Chest CT. The results suggest that Block Attention is particularly effective in medical imaging scenarios where subtle local features play a crucial role in classification.

The combined integration of Grid and Block Attention in the proposed model achieves the highest performance, with accuracies of 0.9954 on IQ-OTH/NCCD and 0.9841 on Chest CT. This significant improvement underscores the complementary nature of these mechanisms, which work synergistically to extract both local and global features, ultimately enhancing the model's robustness and generalizability. This component-wise evaluation confirms the critical roles of Grid and Block Attention mechanisms in the proposed architecture. Their integration not only improves individual feature extraction capabilities but also enables the model to achieve state-of-the-art performance across diverse datasets.

The lightweight design of the proposed model offers significant advantages for real-time use in clinical settings. With only 18.1 million parameters, the model achieves computational efficiency, reducing memory and processing power requirements. This makes it highly suitable for integration in resource-constrained environments, such as portable diagnostic devices or edge computing platforms in remote or underserved areas. In addition, the reduced computational requirements enable faster inference times, which is critical for providing immediate feedback to clinicians during diagnostic workflows. The scalability of the model ensures that it can be implemented on a variety of hardware configurations, from high-performance GPUs to more modest processing units, without compromising accuracy. By optimizing both performance and efficiency, the model bridges the gap between state-of-the-art research and practical application in real-world clinical scenarios, facilitating wider accessibility and adoption. The lightweight design of the proposed model offers significant advantages for real-time use in clinical settings. With only 18.1 million parameters, the model achieves computational efficiency, reducing memory and processing power requirements. This makes it highly suitable for integration in resource-constrained environments, such as portable diagnostic devices or edge computing platforms in remote or underserved areas. In addition, the reduced computational requirements enable faster inference times, which is critical for providing immediate feedback to clinicians during diagnostic workflows. The scalability of the model ensures that it can be implemented on a variety of hardware configurations, from high-performance GPUs to more modest processing units, without compromising accuracy. By optimizing both performance and efficiency, the model bridges the gap between state-of-the-art research and practical application in real-world clinical scenarios, facilitating wider accessibility and adoption.

V. CONCLUSION

This study introduces a DL model with a hybrid approach that combines CNNs and ViTs with improved InceptionNeXt blocks and grid and block attention methods, resulting in a robust solution for lung cancer detection and classification. By utilizing the strengths of these components, the model effectively captures both fine-grained and large-scale features, enabling accurate differentiation of malignant and benign nodules as well as specific cancer subtypes, such as adenocarcinoma, large cell carcinoma, and squamous cell carcinoma. Evaluation on public datasets, Chest CT and IQ-OTH/NCCD, demonstrated the model's exceptional performance, achieving accuracies of 98.41% and 99.54%, respectively. These results highlight its superiority over state-of-the-art methods, while maintaining a lightweight architecture with only 18.1 million parameters. The integration of multi-scale feature processing and attention mechanisms addresses significant gaps in the literature, offering a more comprehensive and efficient diagnostic framework. This model holds promise for enhancing early lung cancer detection, a critical factor in improving patient outcomes and survival rates. Its lightweight design ensures scalability and feasibility for real-world clinical applications. Future efforts will focus on refining the model, extending its applicability to other imaging modalities, and testing its generalizability across diverse populations and datasets to maximize its clinical impact.

DECLARATION OF COMPETING INTEREST

The authors declare no competing interests.

REFERENCES

[1] S. H. Hosseini, R. Monsefi, and S. Shadroo, "Deep learning applications for lung cancer diagnosis: A systematic review," *Multimedia Tools Appl.*, vol. 83, no. 5, pp. 14305–14335, Feb. 2024, doi: 10.1007/s11042-023-16046-w.

[2] A. Agarwal, K. Patni, and D. Rajeswari, "Lung cancer detection and classification based on alexnet CNN," in *Proc. 6th Int. Conf. Commun. Electron. Syst. (ICCES)*, Jul. 2021, pp. 1390–1397, doi: 10.1109/ICCES51350.2021.9489033.

[3] R. L. Siegel, A. N. Giaquinto, and A. Jemal, "Cancer statistics, 2024," *CA, A Cancer J. Clinicians*, vol. 74, no. 1, pp. 12–49, Jan. 2024, doi: 10.3322/caac.21820.

[4] P. Bhowal, S. Sen, J. D. Velasquez, and R. Sarkar, "Fuzzy ensemble of deep learning models using choquet fuzzy integral, coalition game and information theory for breast cancer histology classification," *Expert Syst. Appl.*, vol. 190, Mar. 2022, Art. no. 116167, doi: 10.1016/j.eswa.2021.116167.

[5] N. Maleki, Y. Zeinali, and S. T. A. Niaki, "A k-NN method for lung cancer prognosis with the use of a genetic algorithm for feature selection," *Expert Syst. Appl.*, vol. 164, Feb. 2021, Art. no. 113981, doi: 10.1016/j.eswa.2020.113981.

[6] X. Fu, L. Bi, A. Kumar, M. Fulham, and J. Kim, "An attention-enhanced cross-task network to analyse lung nodule attributes in CT images," *Pattern Recognit.*, vol. 126, Jun. 2022, Art. no. 108576, doi: 10.1016/j.patcog.2022.108576.

[7] S. Qiu, Q. Guo, D. Zhou, Y. Jin, T. Zhou, and Z. He, "Isolated pulmonary nodules characteristics detection based on CT images," *IEEE Access*, vol. 7, pp. 165597–165606, 2019, doi: 10.1109/ACCESS.2019.2951762.

[8] A. E. Celik, J. Rasheed, and A. Yahyaoui, "Machine learning approaches for lung cancer prediction," in *Proc. 12th Int. Conf. Adv. Comput. Inf. Technol. (ACIT)*, 2022, pp. 540–543, doi: 10.1109/ACIT54803.2022.9913114.

[9] D. Riquelme and M. Akhloufi, "Deep learning for lung cancer nodules detection and classification in CT scans," *AI*, vol. 1, no. 1, pp. 28–67, Jan. 2020, doi: 10.3390/ai1010003.

[10] S. Shandilya and S. R. Nayak, "Analysis of lung cancer by using deep neural network," in *Innovation in Electrical Power Engineering, Communication, and Computing Technology: Proceedings of Second IEPCCT 2021*. Singapore: Springer, 2022, pp. 427–436.

[11] A. Atmakuru, S. Chakraborty, O. Faust, M. Salvi, P. Datta Barua, F. Molinari, U. R. Acharya, and N. Homaira, "Deep learning in radiology for lung cancer diagnostics: A systematic review of classification, segmentation, and predictive modeling techniques," *Expert Syst. Appl.*, vol. 255, Dec. 2024, Art. no. 124665, doi: 10.1016/j.eswa.2024.124665.

[12] R. Javed, T. Abbas, A. H. Khan, A. Daud, A. Bukhari, and R. Alharbey, "Deep learning for lungs cancer detection: A review," *Artif. Intell. Rev.*, vol. 57, no. 8, p. 197, Aug. 2024, doi: 10.1007/s10462-024-10807-1.

[13] L. Wang, "Deep learning techniques to diagnose lung cancer," *Cancers*, vol. 14, no. 22, p. 5569, Nov. 2022, doi: 10.3390/cancers14225569.

[14] A. Asuntha and A. Srinivasan, "Deep learning for lung cancer detection and classification," *Multimedia Tools Appl.*, vol. 79, nos. 11–12, pp. 7731–7762, Mar. 2020, doi: 10.1007/S11042-019-08394-3.

[15] A. Halder and D. Dey, "MorphAttnNet: An attention-based morphology framework for lung cancer subtype classification," *Biomed. Signal Process. Control*, vol. 86, Sep. 2023, Art. no. 105149, doi: 10.1016/j.bspc.2023.105149.

[16] M. H. Sabzalian, F. Kharajinezhadian, A. Tajally, R. Reihanisaransari, H. Ali Alkhazaleh, and D. Bokov, "New bidirectional recurrent neural network optimized by improved ebola search optimization algorithm for lung cancer diagnosis," *Biomed. Signal Process. Control*, vol. 84, Jul. 2023, Art. no. 104965, doi: 10.1016/j.bspc.2023.104965.

[17] L. Ma, H. Wu, and P. Samundeeswari, "GoogLeNet-AL: A fully automated adaptive model for lung cancer detection," *Pattern Recognit.*, vol. 155, Nov. 2024, Art. no. 110657, doi: 10.1016/j.patcog.2024.110657.

[18] N. Gautam, A. Basu, and R. Sarkar, "Lung cancer detection from thoracic CT scans using an ensemble of deep learning models," *Neural Comput. Appl.*, vol. 36, no. 5, pp. 2459–2477, Feb. 2024, doi: 10.1007/s00521-023-09130-7.

[19] 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.*, vol. 243, Jan. 2024, Art. no. 107879, doi: 10.1016/j.cmpb.2023.107879.

[20] A. Heidari, D. Javaheri, S. Toumaj, N. J. Navimipour, M. Rezaei, and M. Unal, "A new lung cancer detection method based on the chest CT images using federated learning and blockchain systems," *Artif. Intell. Med.*, vol. 141, Jul. 2023, Art. no. 102572, doi: 10.1016/j.artmed.2023.102572.

[21] A. R. Bushara, R. S. Vinod Kumar, and S. S. Kumar, "An ensemble method for the detection and classification of lung cancer using computed tomography images utilizing a capsule network with visual geometry group," *Biomed. Signal Process. Control*, vol. 85, Aug. 2023, Art. no. 104930, doi: 10.1016/j.bspc.2023.104930.

[22] R. Raza, F. Zulfiqar, M. O. Khan, M. Arif, A. Alvi, M. A. Iftikhar, and T. Alam, "Lung-EffNet: Lung cancer classification using EfficientNet from CT-scan images," *Eng. Appl. Artif. Intell.*, vol. 126, Nov. 2023, Art. no. 106902, doi: 10.1016/j.engappai.2023.106902.

[23] J. Subash and S. Kalaivani, "Dual-stage classification for lung cancer detection and staging using hybrid deep learning techniques," *Neural Comput. Appl.*, vol. 36, no. 14, pp. 8141–8161, May 2024, doi: 10.1007/s00521-024-09425-3.

[24] M. Nahiduzzaman, L. F. Abdulrazak, M. A. Ayari, A. Khandakar, and S. M. R. Islam, "A novel framework for lung cancer classification using lightweight convolutional neural networks and ridge extreme learning machine model with Shapley additive exPlanations (SHAP)," *Expert Syst. Appl.*, vol. 248, Aug. 2024, Art. no. 123392, doi: 10.1016/j.eswa.2024.123392.

[25] Z. He, D. Jia, C. Zhang, Z. Li, and N. Wu, "An automatic darknet-based immunohistochemical scoring system for IL-24 in lung cancer," *Eng. Appl. Artif. Intell.*, vol. 128, Feb. 2024, Art. no. 107485, doi: 10.1016/j.engappai.2023.107485.

[26] J. Gowthamy and S. Ramesh, "A novel hybrid model for lung and colon cancer detection using pre-trained deep learning and KELM," *Expert Syst. Appl.*, vol. 252, Oct. 2024, Art. no. 124114, doi: 10.1016/j.eswa.2024.124114.

[27] T.-O. Tran, T. H. Vo, and N. Q. K. Le, "Omics-based deep learning approaches for lung cancer decision-making and therapeutics development," *Briefings Funct. Genomics*, vol. 23, no. 3, pp. 181–192, May 2024, doi: 10.1093/bfgp/elad031.

[28] M. G. Lanjewar, K. G. Panchbhai, and P. Charanarur, "Lung cancer detection from CT scans using modified DenseNet with feature selection methods and ML classifiers," *Expert Syst. Appl.*, vol. 224, Aug. 2023, Art. no. 119961, doi: 10.1016/j.eswa.2023.119961.

[29] I. Naseer, S. Akram, T. Masood, M. Rashid, and A. Jaffar, "Lung cancer classification using modified U-Net based lobe segmentation and nodule detection," *IEEE Access*, vol. 11, pp. 60279–60291, 2023, doi: 10.1109/ACCESS.2023.3285821.

[30] S. R. Quasar, R. Sharma, A. Mittal, M. Sharma, D. Agarwal, and I. de La Torre Díez, "Ensemble methods for computed tomography scan images to improve lung cancer detection and classification," *Multimedia Tools Appl.*, vol. 83, no. 17, pp. 52867–52897, Nov. 2023, doi: 10.1007/s11042-023-17616-8.

[31] V. Bishnoi and N. Goel, "Tensor-RT-based transfer learning model for lung cancer classification," *J. Digit Imag.*, vol. 36, no. 4, pp. 1364–1375, Aug. 2023, doi: 10.1007/s10278-023-00822-z.

[32] H. Xiao, Q. Liu, and L. Li, "MFMANet: Multi-feature multi-attention network for efficient subtype classification on non-small cell lung cancer CT images," *Biomed. Signal Process. Control*, vol. 84, Jul. 2023, Art. no. 104768, doi: 10.1016/j.bspc.2023.104768.

[33] C. Lin and T.-Y. Yang, "A fusion-based convolutional fuzzy neural network for lung cancer classification," *Int. J. Fuzzy Syst.*, vol. 25, no. 2, pp. 451–467, Mar. 2022, doi: 10.1007/s40815-022-01399-5.

[34] S. K. Lakshmanaprabu, S. N. Mohanty, K. Shankar, N. Arunkumar, 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, doi: 10.1016/j.future.2018.10.009.

[35] R. Mahum and A. S. Al-Salman, "Lung-RetinaNet: Lung cancer detection using a RetinaNet with multi-scale feature fusion and context module," *IEEE Access*, vol. 11, pp. 53850–53861, 2023, doi: 10.1109/ACCESS.2023.3281259.

[36] S. U. Atiya, N. V. K. Ramesh, and B. N. K. Reddy, "Classification of non-small cell lung cancers using deep convolutional neural networks," *Multimedia Tools Appl.*, vol. 83, no. 5, pp. 13261–13290, Jul. 2023, doi: 10.1007/s11042-023-16119-w.

[37] H. Alyasriy, 2020, "The IQ-OTHNCCD lung cancer dataset," vol. 1, doi: 10.17632/BHMDR45BH2.1.

[38] *Chest CT-Scan Images Dataset*. Accessed: Oct. 8, 2024. [Online]. Available: https://www.kaggle.com/datasets/mohamedhanyyy/chest-ctscan-images

[39] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[40] Q. Wang, Y. Liu, Z. Xiong, and Y. Yuan, "Hybrid feature aligned network for salient object detection in optical remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5624915, doi: 10.1109/TGRS.2022.3181062.

[41] I. Pacal, D. Karaboga, A. Basturk, B. Akay, and U. Nalbantoglu, "A comprehensive review of deep learning in colon cancer," *Comput. Biol. Med.*, vol. 126, Nov. 2020, Art. no. 104003, doi: 10.1016/j.compbiomed.2020.104003.

[42] H. Chen, J. Liu, Q.-M. Wen, Z.-Q. Zuo, J.-S. Liu, J. Feng, B.-C. Pang, and D. Xiao, "CytoBrain: Cervical cancer screening system based on deep learning technology," *J. Comput. Sci. Technol.*, vol. 36, no. 2, pp. 347–360, Apr. 2021, doi: 10.1007/s11390-021-0849-3.

[43] K. Han, Y. Wang, H. Chen, X. Chen, J. Guo, Z. Liu, Y. Tang, A. Xiao, C. Xu, Y. Xu, Z. Yang, Y. Zhang, and D. Tao, "A survey on vision transformer," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 1, pp. 87–110, Jan. 2023, doi: 10.1109/TPAMI.2022.3152247.

[44] E. Aslan and Y. Özüpak, "Diagnosis and accurate classification of apple leaf diseases using vision transformers," *Comput. Decis. Making: Int. J.*, vol. 1, pp. 1–12, Jul. 2024, doi: 10.59543/comdem.v1i.10039.

[45] E. Aslan, "Diagnosis of pneumonia from chest X-ray images with vision transformer approach," *Gazi Univ. J. Sci. A, Eng. Innov.*, vol. 11, no. 2, pp. 324–334, Jun. 2024, doi: 10.54287/gujsa.1464311.

[46] I. Pacal, "MaxCerVixT: A novel lightweight vision transformer-based approach for precise cervical cancer detection," *Knowledge-Based Syst.*, vol. 289, Apr. 2024, Art. no. 111482, doi: 10.1016/j.knosys.2024.111482.

[47] Z. Tu, H. Talebi, H. Zhang, F. Yang, P. Milanfar, A. Bovik, and Y. Li, "MaxViT: Multi-axis vision transformer," in *Proc. Eur. Conf. Comput. Vis.*, S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, Eds., Cham, Switzerland: Springer, 2022, pp. 459–479.

[48] I. Pacal, "Enhancing crop productivity and sustainability through disease identification in maize leaves: Exploiting a large dataset with an advanced vision transformer model," *Expert Syst. Appl.*, vol. 238, Mar. 2024, Art. no. 122099, doi: 10.1016/j.eswa.2023.122099.

[49] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and efficient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 10778–10787, doi: 10.1109/CVPR42600.2020.01079.

[50] S. Yıkmiş, M. Türkol, I. Pacal, A. D. Altan, N. Tokatlı, G. Abdi, N. T. Demirok, and R. M. Aadil, "Optimization of bioactive compounds and sensory quality in thermosonicated black carrot juice: A study using response surface methodology, gradient boosting, and fuzzy logic," *Food Chem., X*, vol. 25, Jan. 2025, Art. no. 102096, doi: 10.1016/j.fochx.2024.102096.

[51] B. Ozdemir and I. Pacal, "An innovative deep learning framework for skin cancer detection employing ConvNeXtV2 and focal self-attention mechanisms," *Results Eng.*, vol. 25, Mar. 2025, Art. no. 103692, doi: 10.1016/j.rineng.2024.103692.

[52] W. Yu, P. Zhou, S. Yan, and X. Wang, "InceptionNeXt: When inception meets ConvNeXt," 2024, *arXiv:2303.16900*.

[53] I. Pacal and G. Işık, "Utilizing convolutional neural networks and vision transformers for precise corn leaf disease identification," *Neural Comput. Appl.*, vol. 2024, pp. 1–18, Dec. 2024, doi: 10.1007/s00521-024-10769-z.

[54] E. Aslan and Y. Özüpak, "Detection of road extraction from satellite images with deep learning method," *Cluster Comput.*, vol. 28, no. 1, p. 72, Feb. 2025, doi: 10.1007/s10586-024-04880-y.

[55] E. Aslan and Y. Özüpak, "Classification of blood cells with convolutional neural network model," *Bitlis Eren Üniversitesi Fen Bilimleri Dergisi*, vol. 13, no. 1, pp. 314–326, Mar. 2024, doi: 10.17798/bitlisfen.1401294.

[56] 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," *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 336–359, Feb. 2020, doi: 10.1007/s11263-019-01228-7.

BURHANETTIN OZDEMIR received the Ph.D. degree in educational measurement and statistics from Hacettepe University.

From 2014 to 2015, he was a Visiting Scholar with the University of Illinois at Urbana–Champaign (UIUC). He was a Psychometrician with the National Center for Assessment (NCA, Qiyas), from 2017 to 2019. He is currently an Assistant Professor with the College of Business, Alfaisal University. His research spans a wide range of fields, including computerized and multistage adaptive testing, structural equation modeling (SEM), differential item functioning (DIF), and language testing. His interests also extend to business intelligence, machine learning, artificial intelligence, and the integration of technology into psychometrics. He has published extensively in applied statistics, psychometrics, language testing, and information and communication technologies (ICT), contributing to both theoretical advancements and practical applications in these fields.

EMRAH ASLAN received the bachelor's degree in computer engineering and the bachelor's and master's degrees in electrical and electronics engineering from Harran University, Türkiye, in 2013, 2019, and 2016, respectively, and the Ph.D. degree in electrical and electronics engineering from Dicle University, in 2023.

He is currently an Assistant Professor with Mardin Artuklu University. He has taken an interdisciplinary approach and aims to contribute to both academic and technological progress. His research interests include humanoid robots, deep learning, and renewable energy.

ISHAK PACAL received the B.Sc. degree in computer engineering from Harran University, the M.Sc. degree in electronic communications and computer engineering from the University of Nottingham, and the Ph.D. degree in computer science (artificial intelligence) from Erciyes University, in 2022.

He is currently an Assistant Professor with Igdir University. With a strong academic background and more than 45 publications in SCI-indexed journals, his research interests encompass medical image processing, artificial intelligence in healthcare, and AI-driven applications in agriculture.