Advances in Artificial Intelligence and Machine Learning; Research 6 (2) 5392-5415 Received 05-01-2026; Accepted 21-04-2026; Published online 28-04-2026

A Robust Hybrid Framework for Lung Nodule Detection and Classification from CT Images

B. Usha Priya
Research Scholar, Department of Computer Science Engineering, JNTUA College of Engineering, Ananthapuramu, Andhra Pradesh 515002, India.

V. Lokeswara Reddy
Professor, Department of Computer Science Engineering, KSRM College of Engineering, Andhra Pradesh 516005, India.

Corresponding Author: B. Usha Priya

Abstract

Introduction/Objective: The detection of lung cancer in the early stages has the potential for saving lives and reducing mortality rates. However, the existing methods are unable to jointly capture local features, global relationships, and temporal changes, which are affecting the detection of lung nodules and their classification.

Methods: This study proposes a novel hybrid deep learning framework called NoduleVision-Net, which has the potential for detecting lung cancer in the early stages. The proposed framework combines various modules for local feature extraction, global understanding, and temporal understanding, which are then integrated by the Nodule-Net framework for a novel, robust, and clinically interpretable detection and classification of lung nodules. Once the clean and consistent data are obtained after the preprocessing steps involving Non-Local Means filtering and brightness normalization, the proposed framework combines the modules involving local features, global understanding, and temporal understanding for a novel, robust, and clinically interpretable detection and classification of lung nodules

Results: The performance of the Nodule-Net model was found to be superior in the classification of lung cancer, with 99.98% accuracy, precision, recall, and F1-score achieved in the Chest CT dataset.

Discussion: The proposed model proves its effectiveness in the detection of lung cancer through its accuracy and efficiency in the classification process. Its high precision and reduced complexity outperform prior methods by enhancing early diagnosis, informing clinical decision-making, and advancing automated imaging analysis in medical practice.

Conclusion: This model, with high accuracy, precision, and reduced complexity in lung cancer classification, outperforms traditional methods by enhancing early diagnosis, informing clinical decision-making, and advancing automated imaging analysis in medical practice.

Citation: B. Usha Priya and V. Lokeswara Reddy. A Robust Hybrid Framework for Lung Nodule Detection and Classification from CT Images. **Advances in Artificial Intelligence and Machine Learning**. 2026;6(2):299. https://dx.doi.org/10.54364/AAIML.2026.62299

Keywords: Early lung cancer diagnosis, NoduleVision-Net (or Hybrid Nodule-Net Framework), Medical image analysis, Temporal feature modeling, Explainable Artificial Intelligence (XAI), CNN, Deep learning, Chest CT

1. INTRODUCTION

Lung cancer is often treated to be the cancer-related source of death that occurs the most often globally. Lung cancer tops the list as the malignancy with the highest death rate. It is thus the most fatal type of cancer worldwide [1]. Most instances of lung cancer are attributable to smoking, but a small number of non-smokers also contract this disease. Lung cancer tumors may be located very deep within the lung tissue, and hence they can vary greatly in their locations, intensities, and shapes [2]. While non-small cell lung cancers (NSCLC) make up approximately 85% of all lung cancers, small cell lung cancers represent only about 15% [3, 4]. Biopsies and bronchoscopy are two of the many diagnostic techniques that investigate lung tissue to identify cancer-related cells. A biopsy can predict lung cancer, but maintaining accuracy and precision is difficult. Although screening techniques have helped to forecast lung cancer, maintaining early diagnosis and accuracy in cancer detection is difficult [5, 6]. Consequently, primary detection of carcinoma of the lungs is essential for maximizing survival.

Traditional Computer-aided diagnosis (CAD) systems' effectiveness is determined by the subjective and time-consuming intermediate image processing phases (such as manually extracting morphological and statistical information from a picture) [7]. The most prominent pathological test for diagnosis is the CT scan, which is widely used. It offers a three-dimensional evaluation of the lesion using excellent clarity and high-intensity images of the lung in different postures. In conventional studies, grayscale CT scan pictures were used to create an automated technique for cancer diagnosis [8]. Conventional CT analysis is time-consuming and labor-intensive, and traditional lung cancer detection with CT frequently yields false-positive test findings [9].

Deep learning (DL) has been demonstrated to be very helpful in computational pathology, among other applications. It has been demonstrated that a DL model can successfully identify and predict the mutations from the histology of non-small cell lung cancer [10]. However, until now, there have only been a few deep learning studies that focused on using PET imaging for the finding of lung cancer, and pairing of PET scans and deep learning for the NSCLC subtype classification is the most neglected topic [11]. Due to the low accuracy and dependability that might result from inadequate data, typical deep learning algorithms have sought to analyze big datasets from public sources [12]. CT scan images may be utilized to classify benign and malignant tissues using convolutional neural networks (CNNs). To provide outstanding results, the application of an outstanding performance technology with low expenses and wide applicability could assist the lung cancer prognosis procedure. Utilizing pictures from CT scans, the therapeutic imaging procedure was created to identify lung cancer [13].

In the traditional approach, it is possible to make a distinction among the various kinds of lung cancer in addition to detecting skeletal metastatic tumors in lung cancer patients. As a result, it is more challenging than the current multi-disease categorization. Furthermore, the original tumors from which the bone metastases originated have not previously been identified [14]. To analyze CT lung cancer detection images and assess lung functions in lung cancer patients, Mask-RCNN,

an optimization-based technique, was used for automated segmentation. Additionally, serum tumor detection was used to evaluate the consequences of recognizing characteristics of benign and malignant cancers [15]. To the knowledge, no research has used the segmentation approach to identify lung cancer that has been pathologically confirmed on chest radiographs [16]. Fast, precise, and mainly automated lung cancer detection is essential in today's developing new technologies and is far more crucial than using outdated techniques. Anomalies in medical imaging are frequently detected and treated using conventional methods. When using radiology CT scans to diagnose lung nodules, CAD has grown in importance [17]. By analyzing the traditional methods diagnosis of lung cancer is still challenging; hence, it is essential to explore a novel framework to diagnose and classify lung cancer.

Some of the recent studies in lung cancer detection, including cloud-based CNNs, hybrid models, traditional approaches like PET/CT analysis, and ResNet-LSTM, are discussed in the following section. Kasinathan et al. (2022) [18], came up with a method of identification and confirmation of various phases of lung tumor development; besides that, they introduced a deep neural model and a cloud-based data acquisition scheme for the classification of pulmonary disorder phases. The technique suggested by the authors is a hybrid method for PET/CT images, being a Cloud-based Lung Tumour Detector and Stage Classifier (Cloud, LTDSC). First, the prompting of an active contour model for lung tumor segmentation in the proposed Cloud, LTDSC, was followed by the design and testing of a multilayer convolutional neural network (M-CNN) to classify different phases of lung cancer by the usage of benchmark images. In one case of cloud computing due to memory limitations, the proposed cloud system is only capable of analyzing the records of 65 users at one time, which is not very efficient for large hospital setups.

Ragab et al. (2022) [19], offered the DHODCNNLCC model, which uses deep convolutional neural networks for recognizing and categorizing lung cancer. The first pre-processing steps for the proposed DHODCNN-LCC approach are contrast enhancement and noise reduction. In addition, the Nadam optimizer with the RefineDet model is used to extract the features from the pre-processed pictures. For the categorization of lung nodules, the denoising stacked autoencoder (DSAE) model is also used. Finally, higher classification performance is achieved by optimally improving the DSAE model's hyperparameters using the deer hunting optimization algorithm (DHOA). However, the performance of the classification is not improved by the use of image segmentation algorithms.

Wang et al. (2022) [20], developed a two-stage diagnosis method with Dynamic R, CNN, or a dynamic region-based CNN. This method is capable of boosting the accuracy of medical diagnosis and is marked by a high performance level in lung cancer identification. It can, in some cases, prevent erroneous and missed detection. However, it cannot increase the network's precision and does not demonstrate the model's widespread application in cancer detection.

Zeyu et al. (2022) [21], presented the LCGANT hybrid framework. Our framework consists of two essential components in particular. To create artificial images of lung cancer, the first phase uses a deep convolutional GAN (LCGAN). The second part deals with categorising lung cancer images in three classes by employing the regularization augmented transfer learning model. The presented framework is capable of handling the problem of overfitting for various tasks, such as lung cancer classification. The LCGAN produces synthetic images that differ somewhat from actual images. The generator creates graphics with a 64 x 64-pixel size. For the biomedical field, this is insufficient. High-resolution images are crucial in this situation.

Wankhade et al. (2023) [22], proposed a novel method coupled with cancer early detection through a hybrid neural network (HNN) technique named Cancer Cell Detection (CCDC). Computed tomography (CT) images are analyzed with deep neural networks to obtain the features of the cells in the images. A crucial point is that accurate feature extraction contributes greatly to the early-stage detection of cancer cells, which, in turn, can save the life of the patient. Besides that, this work uses an advanced 3D CNN to lift the recognition accuracy to a higher level. In addition, the proposed method can successfully classify the tumors as either benign or malignant. These models are really sophisticated, and include a substantial quantity of medical photos; it can be predicted that the proposed method would need a considerable amount of processing resources. It is crucial to remember that methods like parallelization, transfer learning, and pruning may be used to decrease the time complexity of a deep neural network.

Cao et al. (2023)[23] came up with a segmenting method based on DL. The view aggregation of the method particularly employed a pixel-wise addition to emphasize the radiopharmaceuticals and high uptake areas. The segmentation of lesions produced images that, along with the viewing process, formed a composite prediction. Conceptually, the model is an encoder, decoder, type, end-to-end network segmentation made up of two sub-networks. Therefore, the feature extraction sub-network can learn the hierarchical characteristics of bone scan images.

Tashtoush et al. (2023) [24], identified DL as a method capable of separating non-small cell lung cancer subtypes, adenocarcinoma, large cell carcinoma, and squamous cell carcinoma, from lung CT pictures. The study cited the use of VGG16, standard CNN, and VGG19. Besides, it came up with a CNN variant that has convolutional block attention modules (CBAM) incorporated in it too. CBAM combines cross-channel and spatial information to feature extraction more effectively. Furthermore, it proposes changes to VGG16 and VGG19 that apply an SVM instead of SoftMax in the classification layer. However, it has not improved learning and improved precision in determining the kind of lung cancer.

Zhang Y (2023) [25], recommended employing NASNet-Large as an encoder in a lung image segmentation model, followed by an architecture for a decoder. In terms of DL models for image segmentation, NASNet-Large is one of the most often used. The segmentation map may be recovered using the suggested NASNet-Large-decoder architecture, which can also enlarge the feature map and retrieve high-level information. It provided a post-processing layer to eliminate the pointless area of the segmentation map to further enhance the segmentation results. Due to its size and the number of parameters it includes, this model is extremely complicated and computationally costly. Because of its intricacy, training, and inference may be time- and resource-intensive.

Although computer-aided diagnosis (CAD) systems have improved significantly in recent years, they still fall short of what is needed in real-world clinical practice. The current models used to analyze CT scans have a significant issue which leads to missed nodules or misclassified cases. While most of the existing CNN-based model are highly capable at identifying local details of CT scans such as fine edges of nodules, they repeatedly miss out on relating to the other parts of a lung to one another. On the other hand, the use of transformer models do provide excellent global relation detection, but require large datasets and do not detect detailed features of a CT scan at the fine level of detail that physicians require. Therefore, solutions that will lead to a hybrid model that is able to utilize the strengths of each of these technologies in a reliable manner will be required to develop better methods for diagnosing lung cancer.

The contribution was highlighted as follows:

- A novel integration of CNNs and transformer-based blocks is proposed to jointly learn detailed local features and long-range global dependencies, avoiding the representational limitations of each individual model.
- Temporal learning mechanisms are incorporated within the framework to effectively characterize lung nodules in terms of longitudinal changes, thereby improving the ability to characterize growth patterns that are significant in early-stage cancer detection.
- A reliable preprocessing technique is utilized that combines Non-Local Means (NLM) filtering with brightness normalization to improve signal quality, reduce noise, and ensure homogeneity within input images.
- Nodule-Net proposes an effective strategy to integrate spatial (local), contextual (global), and temporal features, thereby creating a cohesive space that improves robustness in classification

The rest of this article is organized as follows: Section 2 explains the proposed approach; Section 3 describes the experimental results, major findings, and comparison; and finally Section 4 concludes the research with future research directions.

2. MATERIAL AND METHODS

Nodule-Net is a hybrid DL model designed and developed with the aim of achieving early and accurate detection of lung cancer using images obtained from CT scans. All previous approaches have been limited in that they have concentrated on one feature at the expense of the other. This model seeks to bring the two aspects, local and global, together in a unique workflow. In the preprocessing steps, a unique approach is proposed in that it is all-encompassing. In the preprocessing steps, resampling, normalization, contrast enhancement using CLAHE, Non-Local Means filter, and correction of brightness are included. Standardization improves robustness to scanners as well as visualization of small nodules that might not be captured during routine analysis. The basic idea behind this framework is NoduleVisionNet, which utilizes residual-based hierarchical transformations that enhance local feature extraction capabilities to effectively identify small, irregular, complex textural nodules that are essential in early lung cancer detection. Globally advanced context modeling has been added to this framework, utilizing long-range spatial dependencies across various CT image slices, which is essential in identifying multi-nodule patterns, diffuse malignancies, as well as spatial heterogeneity in lung tissue. The hybrid integration of all these capabilities is what adds another dimension of reliability to the classification results, as opposed to CNN or transformer-based models. The second innovation in Nodule-Net is that it has been designed to track nodules over time, thereby enabling a visualization that can be used to monitor treatment, a consideration that is rarely given any attention. The overall flow diagram of this proposed work is as shown in FIGURE 1.

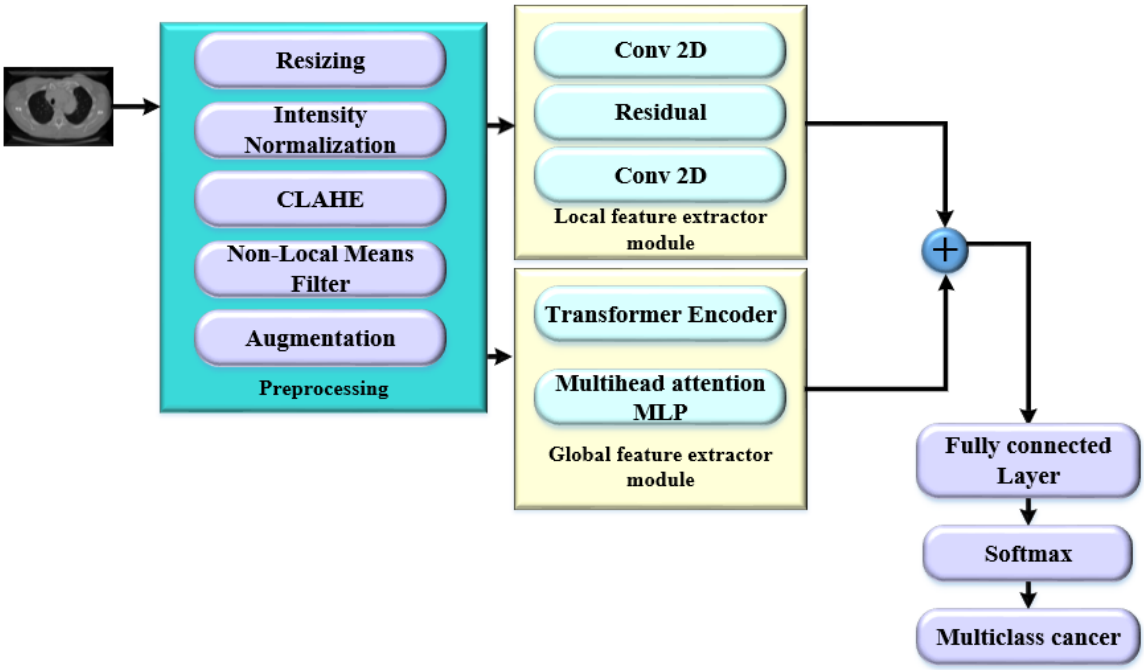


Figure 1: Block diagram for the proposed Nodule-Net Framework

2.1 Preprocessing

The size of the CT scan images is trimmed to a standard size initially, which ensure that the data is of the same size, which is crucial when training the model. For 2D slices, each image I_{orig} of size $H_0 \times W_0$ is resized to a standard dimension $H_r \times W_r$ using bilinear interpolation as equation (1)

$$I_{resized}(x', y') = \sum_{i,j} I_{orig}(i, j) \cdot (1 - |x' - i|) \cdot (1 - |y' - j|) \tag{1}$$

where (x', y') These are coordinates in the resized image.

After spatial standardization, intensity normalization is applied to focus on the lung tissue. The voxel intensities are first clipped to the lung tissue range of -1000 to 400 Hounsfield Units (HU) as equation (2):

$$I_{Clipped} = Clip(I_{resized}, -1000, 400) \tag{2}$$

Then, the min-max normalization is used to scale the clipped values to the [0,1] range as equation (3):

$$I_{norm} = \frac{I_{Clipped} + 1000}{1400} \tag{3}$$

This normalization reduces variability caused by different scanners, enhances the visibility of lung structures, and improves the stability and convergence of DL models during training.

To enhance subtle lung structures, **CLAHE** is applied, which adjusts local contrast while preventing over-amplification. For a pixel in a tile t, the output intensity is given as equation (4).

$$I_{CLAHE}(x, y) = CDF_t(I_{norm}(x, y)) \cdot (I_{max} - I_{min}) + I_{min} \tag{4}$$

where CDF_t is the cumulative distribution function of the histogram in tile t, and I_{min} and I_{max} Are the intensity bounds.

To eliminate noise while retaining detailed structural information, **Non-Local Means** (NLM) filtering is used. And which is performed by the following equation (5):

$$I_{NLM}(v_i) = \sum_{v_j \in \mathcal{N}(v_i)} w(v_i, v_j) I(v_j) \tag{5}$$

with weights:

$$w(v_i, v_j) = \frac{1}{Z(v_i)} exp\left(-\frac{\left\|I(\mathcal{P}_i) - I(\mathcal{P}_j)\right\|_2^2}{h^2}\right) \tag{6}$$

where $\mathcal{P}_i$, $\mathcal{P}_j$ are patches around voxels v_i, v_j respectively, h is the filtering parameter, and $Z(v_i)$ is the normalization factor.

Data augmentation is the final step, which helps improve the overall variability of the data, hence improving the overall ability of the model to generalize well. In the case of 2D slices, the data is rotated $I_{rot} = R_\theta(I_{NLM})$, brightness adjustment $I_{bright} = \alpha I_{NLM} + \beta$.

The final preprocessed and augmented datasets are standardized in space, normalized in intensity, denoised, and contrast-enhanced, which helps in maintaining the quality of the data, thereby improving the representation of the data. These preprocessed data are well-suited for the input of the deep learning algorithm, NoduleVisionNet, that helps in the detection of lung cancer with high accuracy.

2.2 NoduleVisionNet

The conventional CNNs are not effective in detecting fine textures of nodules, small irregular nodules, and dependencies, while conventional models are incapable of capturing long-range dependencies or global spatial relationships, which are crucial for detecting multi-nodule patterns and diffuse malignancies. Additionally, these models are incapable of combining fine local details with contextual information, thus reducing the accuracy in diagnosis. Most importantly, the inability of these methods to detect the growth of nodules in patients has hindered the detection of dangerous cancers and the assessment of the effectiveness of various treatments. To address these limitations, NoduleVisionNet is proposed. The feature extraction process includes the following layers:

- Convolutional Layer: Applies filters to detect low-level features such as edges, shapes, and textures.
- Rectified Linear Unit (ReLU): Serves as a nonlinear activation function, effectively mitigating vanishing gradients while ensuring only positive activations contribute to feature learning.

- Pooling Layer: Reduces spatial dimensions to preserve dominant features while improving generalization.
- Dense Layer: Fully connected neurons integrate extracted features into a comprehensive representation.

For an initial feature extractor to achieve effective local representation learning, a hierarchical residual-based architecture was used as the first layer. The lung CT images input will be treated mathematically according to equation (7).

$$I \in \mathrm{R}^{H \times W \times D} \tag{7}$$

Where the symbols representing height, width, and depth of the image are H, W, and D, respectively. The aim here is to transform I into discriminative feature representations F, such that the features are able to encompass detailed nodule features as well as the global structure. The first stage involves applying a series of convolutional filters to extract low-level image features such as edges, shapes, and nodule boundaries as shown in equation (8).

$$F_1^{(k)} = \sigma(I * W^{(k)} + b^{(k)}) \tag{8}$$

Where, $W^{(k)}$ denotes the convolutional kernel for the k^{th} filter, $b^{(k)}$ denotes the bias term, * denotes convolution operation, and $\sigma(\cdot)$ denotes the ReLU activation function given by $\sigma\,(x) = \max(0, x)$. This non-linear transformation prevents feature collapse and allows for the detection of nodules with irregular boundaries. Instead of using a single convolutional path, this architecture follows a 'split transform merge' principle. In this principle, first, dimension reduction is performed through convolution, followed by several transform operations, and then outputs from all these transform operations are summed. This is called 'cardinality,' and it improves the diversity and richness of features without increasing the depth or complexity of the model. The transformation process can be represented as follows:

$$F\,(x) = x + \sum_{i=1}^{C} T_i(x) \tag{9}$$

where x represents the input, $T_i(x)$ represents the i^{th} transformation function in parallel, and C represents the number of parallel paths. The aggregation of various transformation functions allows more diverse discriminative features to be learned by the network, as opposed to conventional residual connections. The structure of each transformation path is based on a bottleneck structure as described by equation (10)

$$T_i\,(x) = (W_{3\times3}^{(i)} * \phi\left(W_{1\times1}^{(i)} * x\right)) \tag{10}$$

The non-linear activation function is denoted by $\phi\,(\cdot)$ Because of this structure of hierarchical feature capture, it allows for the hierarchical representation of texture, i.e., small nodules to complete diffuse malignancies. After transforming nodule texture, the local representation will become the Equation (11)

$$F_{local} = \mathcal{R}(I) \tag{11}$$

This is used as input for the Hierarchical Residual Operator model $\mathcal{R}(I)$. And this captures various nodule pattern elements across multiple scales, where this method has been used for detecting:

- Small nodules: due to fine-scale convolution filters.
- Texture variations: via residual aggregation.
- Irregular margins: preserved through identity mappings.

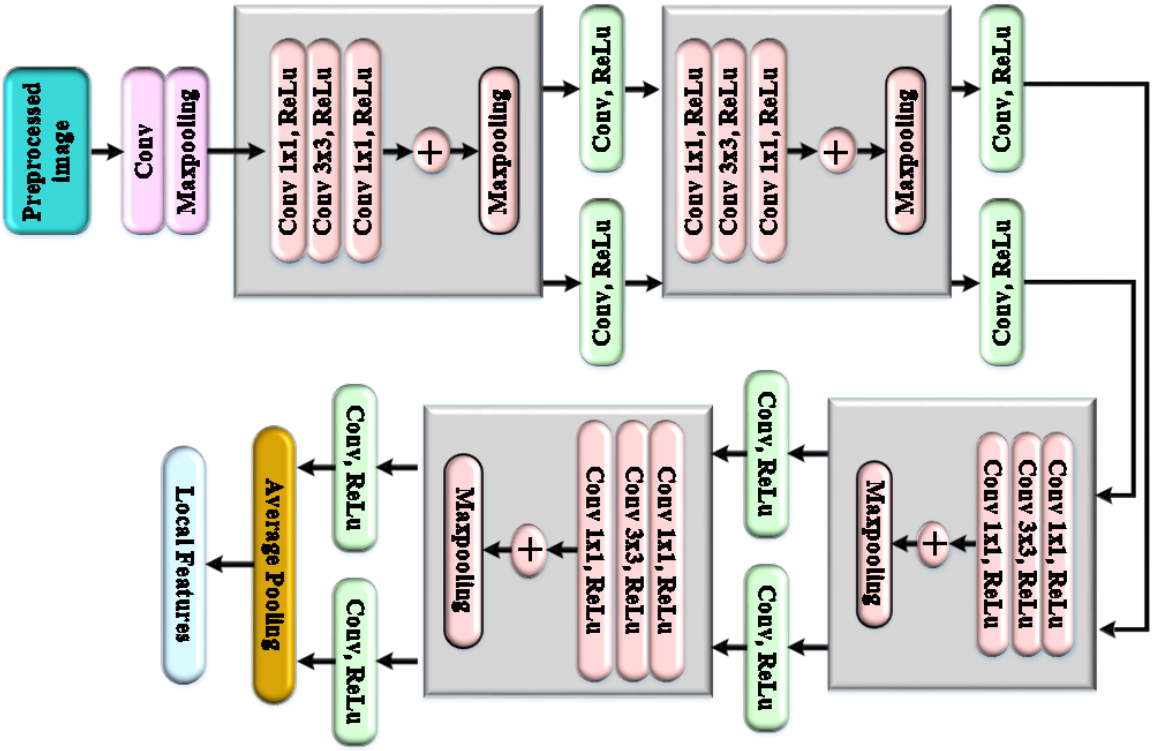


Figure 2: Local Feature Extraction Module

Using this structure, local feature enhancement was achieved as demonstrated in FIGURE 2. The nodular texture, irregular boundaries, and small pulmonary nodules were effectively captured using hierarchical transformations based on residual learning. This enhancement improved the differentiation between benign and malignant nodules, thereby addressing one of the main problems in early lung cancer detection, as nodular features are critical in this case. The local features F_local carry critical information on nodular features, but they are not aware of global spatial relationships within the lung field. For example, nodules may be distributed in various lobes, or malignancy may be diffuse. In order to solve this problem, High-resolution images can be processed effectively with the help of a strong DL model by considering long-range dependencies that represent global context information, as the structure of global feature extraction is represented in FIGURE 3. In contrast to conventional CNNs that mostly focus on local feature extraction, this method is based on a hierarchical structure that allows local features as well as global features to be learned simultaneously.

The process starts with the division of the input image into patches and hierarchical analysis using the two-stage framework. Within each patch, local structures are learned using self-attention

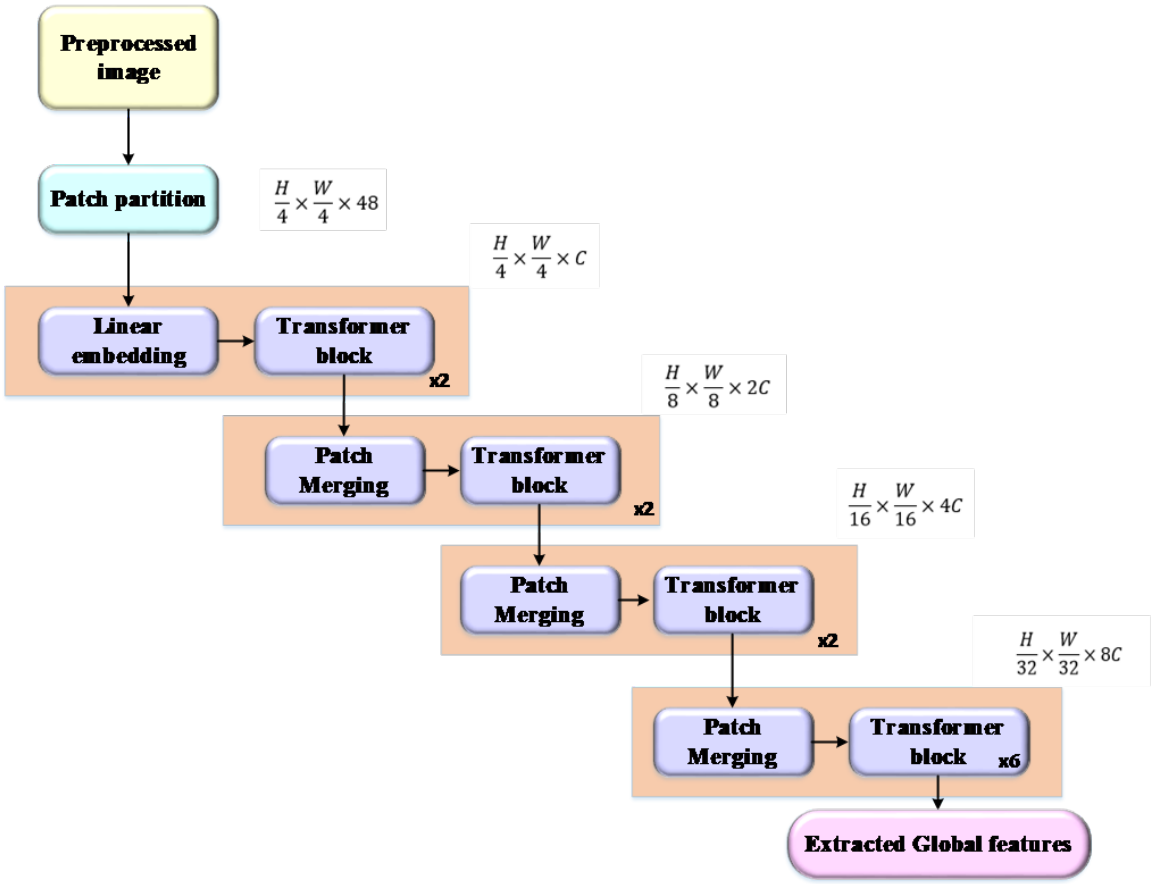


Figure 3: Global feature Extraction Module

mechanisms in the first stage, whereas global dependencies are learned using between-patch self-attention mechanisms in the second stage. To increase the representational power of the network, the concept of shifted windows is incorporated, which enables the network to look at the entire image from different regions. This is very crucial in the joint analysis of microscopic and macroscopic structures within the images. The network architecture is based on the Transformer network with multiple levels of blocks consisting of two key components:

- Multi-Head Self-Attention (MSA) – It primarily focuses on the identification of the relevant contextual information by paying attention to the dependencies across the regions.
- Feedforward Neural Networks (MLP): Enables the learning of complex non-linear data representations, refining feature extraction.

FIGURE 4 illustrates the hierarchical process that switches between window-based multi-head self-attention (W-MSA) and shifted window multi-head self-attention (SW-MSA). This alternation facilitates the modeling of both localized details and broader global structures. The model incorporates

contextual information by sequentially combining it across patches to allow growth and efficiency, and achieves remarkable performance on tasks such as object recognition and image classification.

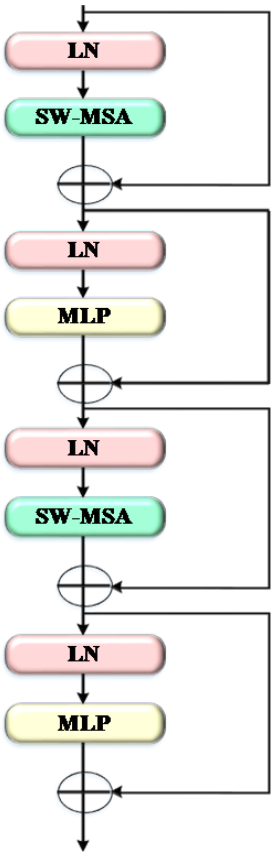


Figure 4: Hierarchical mechanism of SW-MSA for Local–Global Feature Learning

Non-overlapping patches are created by rearranging the feature map and passed through a hierarchical attention mechanism. The local feature map is partitioned into fixed-size patches, given as equation (12)

$$F_{local} \in \mathrm{R}^{H\times W\times d} \rightarrow Z_0 = \{z_1, z_2, \ldots\ldots, z_N\}, \quad z_1 \in \mathrm{R}^d \tag{12}$$

Where $N = \frac{H\times W}{M^2}$ and M denotes the patch size. Each patch is linearly projected to create a d-dimensional embedding. Rather than applying global self-attention (which is expensive for large CT scans), attention is first determined within each local window. When the window contains tokens $Z_w = \{z_1, z_2, \ldots\ldots, z_m\}$, Attention is defined as equation (13)

$$Attention\,(Q,\ K,\ V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \tag{13}$$

with $Q = Z_w W_Q$, $K = Z_w W_K$, $V = Z_w W_V$ where $W_Q, W_K,\ W_V \in \mathrm{R}^{d\times d_k}$. These are learnable projection matrices. This ensures that dependencies within each patch window are modeled effectively, focusing on local contextual details. To extend the model's field of view beyond a single window, the attention windows are shifted by $\frac{M}{2}$ in the next layer. This process allows for interaction between adjacent windows, ensuring that the overall context is taken into account without incurring quadratic complexity.

Mathematically, if W is the window partitioning function, the shifted attention is applied as follows:

$$Z' = SW - MSA\,(Z) = \bigoplus_{w=1}^{N/M^2} Attention\,(Q_w,\ K_w,\ V_w) \tag{14}$$

where $\bigoplus$ denotes the concatenation of the output of the windows, and Q_w, K_w, V_w are computed on the windows shifted as described above. By alternating the operations of W-MSA and SW-MSA, the network builds relationships that cover both local nodules and distant structures in the lung field. After the attention operation, the output of the token is fed into an MLP with residual connections as described in equation (15)

$$Z_{out} = Z' + MLP(LN\,(Z')) \tag{15}$$

where LN is Layer Normalization. This ensures stable training as well as the non-linear refinement of the global features. The hierarchical representation is given by the following equation:

$$F_{global} = \mathcal{A}(F_{local}) \tag{16}$$

where $\mathcal{A}(\cdot)$ is the alternating sequence of W-MSA and SW-MSA layers. F_{global} includes fine-grained features and long-range spatial features, which are essential in robustly discriminating between malignant and benign nodules, even if their spatial locations are distributed over different areas of the lung.

Subsequent to the extraction of local features, i.e., F_{local} which includes nodule textures, boundaries, and fine details, and global features, i.e., F_{global}, which includes long-range dependencies and contextual relationships in the lung field, these features are fused together in the form of equation (17) as given below:

$$F_{fused} = \phi(\alpha \cdot F_{local} \bigoplus \beta \cdot F_{global}) \tag{17}$$

Where $\bigoplus$ represents the concatenation or addition of the two features depending on the fusion strategy and α, β represent the learnable weights that determine the relative importance of local and global features and $\phi(\cdot)$ represents the non-linear transformation, e.g., ReLU or normalization. The fused feature map is flattened and passed through one or more fully connected (dense) layers to project it into a lower-dimensional decision space, as given in equation (18)

$$h = f(W_f \cdot vec\,(F_{fused}) + b_f \tag{18}$$

where $vec\,(F_{fused})$ flattens the fused feature tensor into a vector, W_f, b_f are learnable weights and bias of the fully connected layer, $f(\cdot)$ is a non-linear activation function such as ReLU, h is the hidden feature representation used for classification.

The final hidden representation h is mapped to class logits through another linear transformation:

$$z = W_c h + b_c \tag{19}$$

where W_c, b_c are classification weights and biases, and $z \in \mathrm{R}^3$ corresponds to the multi-class categories (such as Normal, Benign, Malignant). The probability distribution across classes is then obtained via Softmax as equation (20):

$$P\,(y = k\,|x) = \frac{\exp(z_f)}{\sum_{j=1}^{3} \exp(z_j)}, \quad k \in \{1, 2, 3\} \tag{20}$$

This ensures that the outputs are normalized probabilities summing to 1, making the prediction interpretable in clinical terms.

The predicted class $\hat{y}$ is determined by the maximum probability as in equation (21):

$$\hat{y} = \underset{k\in\{1,2,3\}}{\arg\max} P\left(y = k \,|x\right). \tag{21}$$

In both datasets, all of the CT scans are classified into multiple categories, allowing for a complete multiclass evaluation approach. The NoduleVisionNet combines improved extraction of locally relevant features with models that account for the global context, creating a stronger tool for detecting lung cancer than what can be obtained from either CNN or transformer models. The proposed method identifies both fine-grained texture characteristics of nodules as well as long-range spatial relationships between different nodules, so it will address some deficiencies of other methods that were previously used to identify both subtle nodules and how many different types of nodules may occur in a given patient. The use of this approach will allow it to generalize well to multiple types of datasets, making it a valid, interpretable and clinically useful means of identifying and classifying multiple types of CT-based nodules to detect lung cancer early.

3. RESULTS

This section outlines the experimental design, the different properties of the dataset, and the parameters that were configured for the test of the lung cancer detection model. The performance of the system has been measured in terms of diagnostic accuracy, diagnostic precision, and diagnostic stability using conventional metrics such that accuracy, precision, recall, F1 score, and AUC are assessed. The experiment was conducted on a Windows 10 (64 bit) operating system with an Intel Core i5 processor and 16 GB of memory which allowed for a reliable and efficient computing environment. Python 3.10 served as the primary programming language with an array of options for scientific and medical imaging libraries along with support for flexibility when developing. The framework for detecting lung cancer based on the proposed DL approach was developed and trained using TensorFlow and Keras. The framework was also made scalable using optimized performance. The pre-processing of the data, including feature extraction, CT image normalization, and numerical calculations, was implemented using Pandas and NumPy. In addition, parallel machine learning tools and performance measurement were implemented using Scikit-learn. TABLE 1 illustrates the parameters used for simulating the proposed approach.

3.1 Dataset Description

The Chest CT-Scan Images dataset [26] is a well-structured publicly available dataset commonly used for lung cancer classification problems. The total number of images in the Chest CT-Scan Images dataset is approximately 1,000, and the total size of the images in the dataset is approximately 124.96 MB. The total number of images in the dataset is composed of 988 PNG images and 12 JPG images, which can be used to train deep learning models. The images in the Chest CT-Scan Images dataset are systematically classified into four classes, such as adenocarcinoma, large cell carcinoma, squamous cell carcinoma, and normal lung tissues. This diversity in classes helps to learn more general features from the images due to the variety of pathological images. The training

Table 1: Simulation parameters

Parameter	Value
Dataset Split	70:20:10
Optimizer	Adam
Batch size	32
Learning rate	0.001
Loss function	Categorical cross entropy
Number of Epochs	25
Activation Function	ReLU
Dropout Rate	0.5
Model Architecture	Nodule Vision Net

set of the dataset is used to optimize the parameters of the network, the validation set of the dataset is used to select the model and control overfitting, and the test set of the dataset is used to evaluate the generalization of the model. Images are resized to 256 x 256 pixel resolution and median-filtered to reduce noise, allowing for features to be efficiently extracted before training of the model occurs. Additionally, preprocessing for normalization is performed using methods that conform to standard preprocessing for convolutional neural networks. The equal distribution of the various cancer subtypes and the inclusion of patient controls allows for use of this dataset in the development of automated diagnostic systems.

3.2 Visual Analysis

This section presents the experiment outcomes for the proposed framework in the context of detecting lung cancer. The preprocessing technique was used to normalize the images of the CT scan in order to make them suitable for training the model.

The preprocessing pipeline for the Chest CT dataset is depicted in FIGURE 5. The raw CT images are standardized, and then the images are subjected to clipping and normalization to emphasize the lung tissues. Contrast enhancement techniques like CLAHE and denoising using Non-Local Means are employed to enhance the visualization of anatomical features, and augmentation is used to introduce randomness in the images for training the models effectively.

FIGURE 6 shows the visual representation of the Feature correlation heatmap – Chest CT. From the correlation heatmap of the Chest CT features, we can observe that there is a significant correlation between features such as Texture, Contrast, and Shape. These are structural and intensity-related features of the lung tissues. Additionally, we can observe that there is a positive correlation between the contrast and texture features, which are related to highlighting the irregularities in the nodules. Entropy is weakly correlated with the other features.

FIGURE 7 illustrates the performance of the proposed NoduleVisionNet on the Chest CT dataset in terms of accuracy and loss functions for both the training and testing datasets. As depicted in FIGURE 7a, the accuracy for the training and validation sets is observed to increase smoothly and plateau at a very high value of 0.9998. FIGURE 7b illustrates the loss functions for the proposed

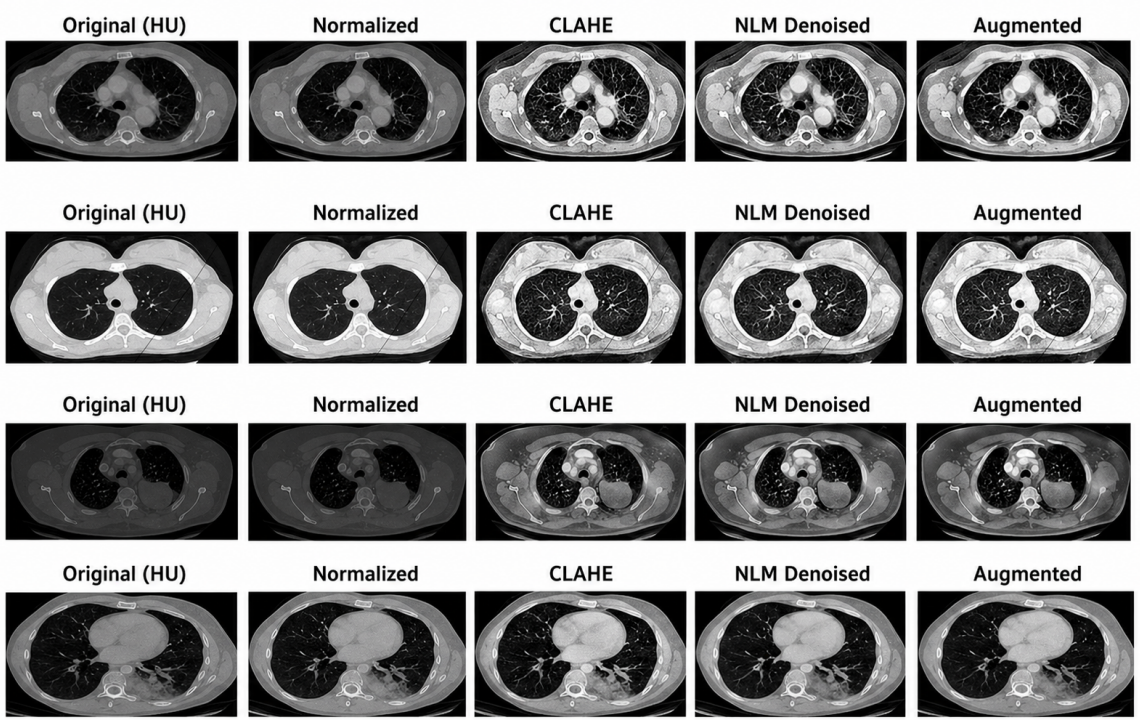


Figure 5: Preprocessed output for Chest CT images dataset

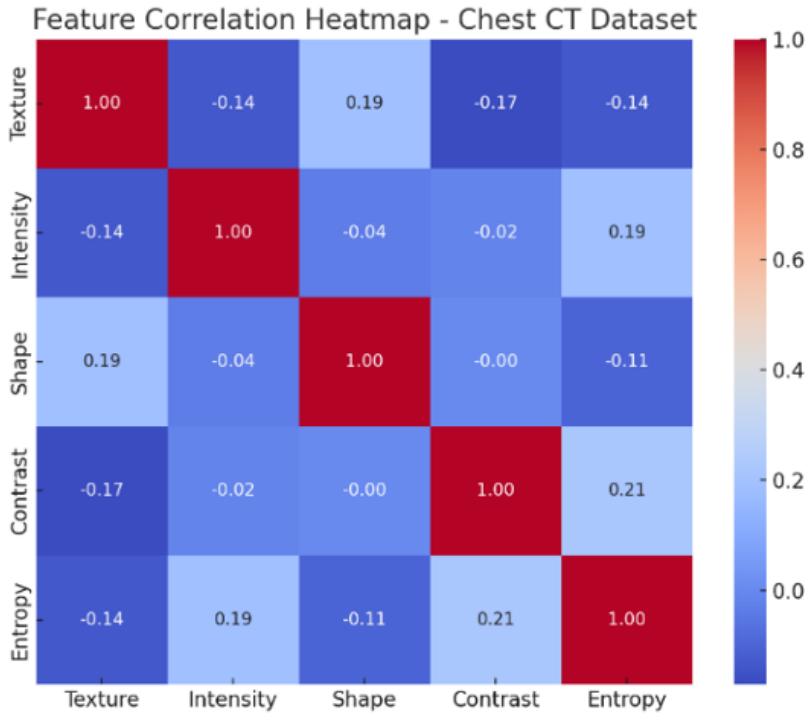
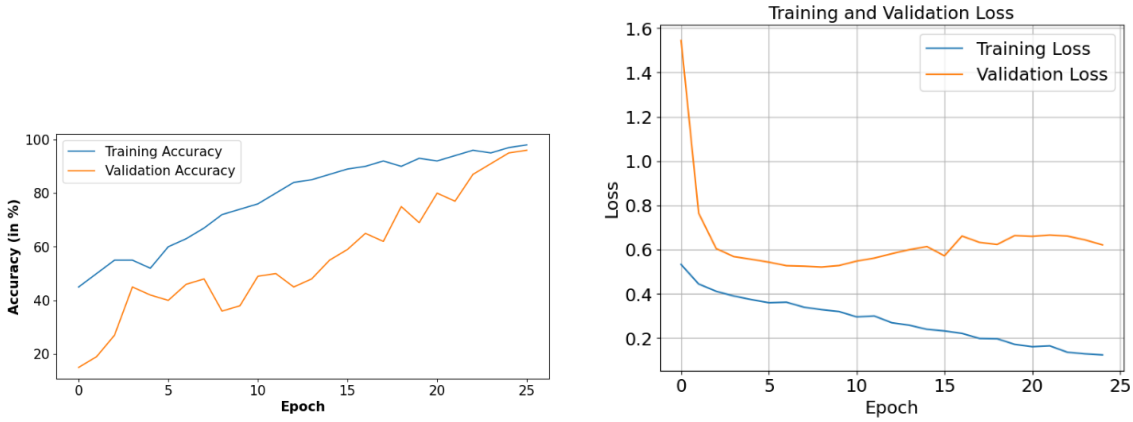


Figure 6: Feature Correlation heatmap



(a) (b)

Figure 7: (a) Accuracy and (b) Graph of Proposed Framework

NoduleVisionNet on the training and validation datasets, in which the loss for the training set is observed to decrease smoothly and plateau at a value of 0.15, while the validation loss plateaus at a value of 0.62 without any signs of divergence.

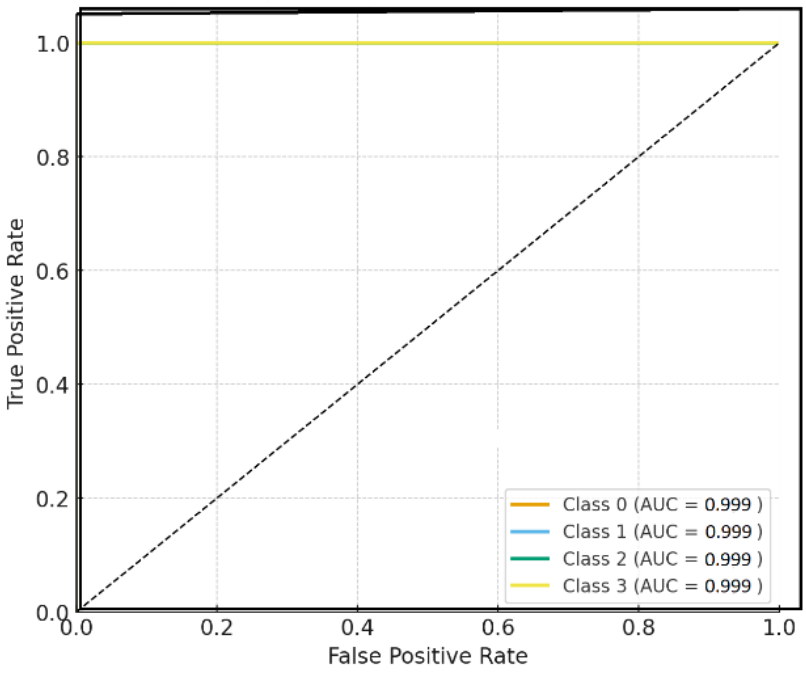


Figure 8: ROC analysis using the Chest CT dataset

The analysis of the ROC curve also validates the accuracy and reliability of the model with respect to the chest CT data set, as presented in FIGURE 8. All the classes were able to achieve an AUC score of 0.999. In other words, the model is highly accurate and reliable with respect to the chest

CT data set. In particular, the model's performance with respect to the Large Cell Carcinoma, Adenocarcinoma, Squamous Cell Carcinoma, and Normal classes is such that the ROC curves are almost at the upper left corner of the plot. This is an indication that the model is highly sensitive and specific in detecting the cases of lung cancer. In other words, the framework's accuracy and reliability are such that it is a sure contender in actual cases of lung cancer detection.

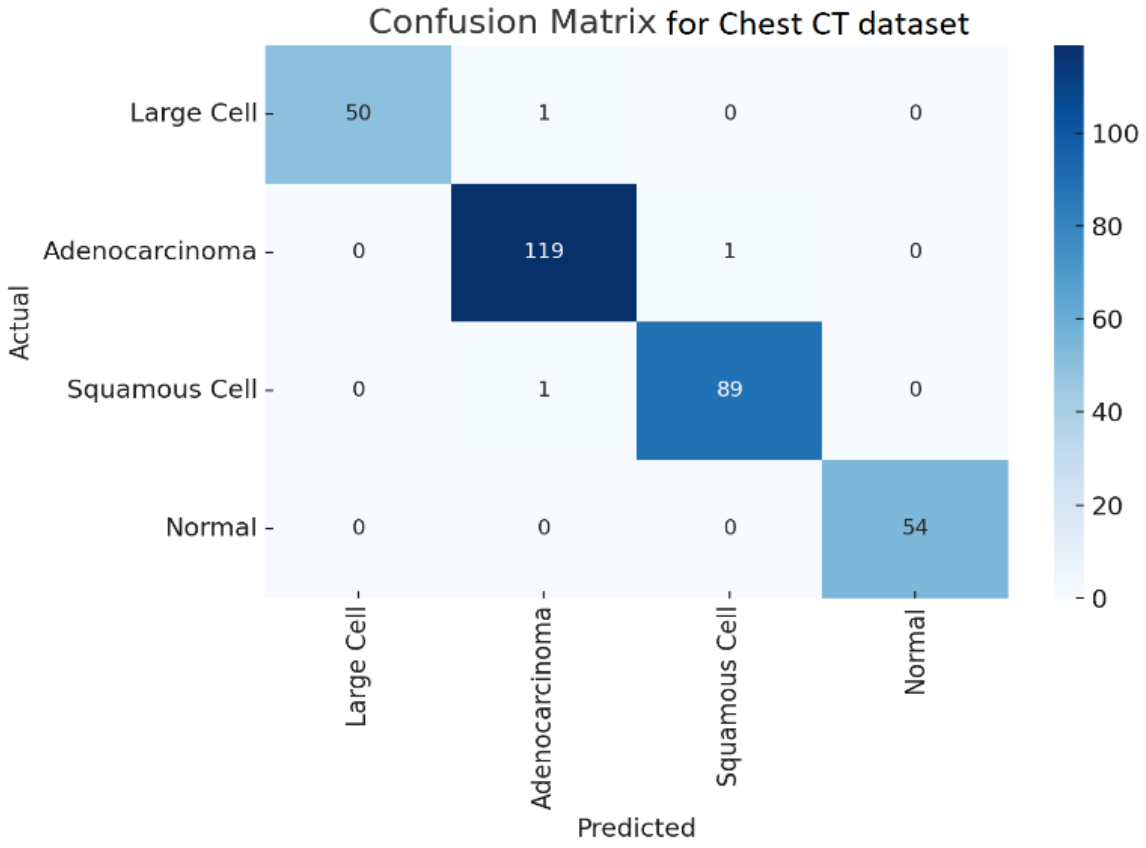


Figure 9: Confusion matrix

The Confusion matrix for the given Chest CT dataset shows that all four classes, namely Large Cell Carcinoma with 51 images, Adenocarcinoma with 120 images, Squamous Cell Carcinoma with 90 images, and Normal with 54 images, are correctly classified by the proposed Nodule-Net model, as shown in FIGURE 9. The classification results showed that all 51 images were correctly classified as Large Cell Carcinoma, all 120 images were correctly classified as Adenocarcinoma, all 90 images were correctly classified as Squamous Cell Carcinoma, and all 54 images were correctly classified as Normal. There were very few instances of misclassification, such as one image of Large cell carcinoma classified as Adenocarcinoma and one image of Squamous Cell Carcinoma classified as Adenocarcinoma. The accuracy is very high, showing that the proposed Nodule-Net is robust enough to classify healthy scans as well as different kinds of lung cancer.

3.3 Qualitative Performance Analysis

In order to ensure the robustness and generalization ability of the proposed Nodule-Net framework, a 5-fold cross-validation approach has been used. In this approach, the dataset has been divided into five equal parts. For each iteration, four parts of the dataset have been used for training, and one part has been used for testing. This process has been repeated for five iterations, so that each part of the dataset has been used for testing exactly once. The performance of the proposed approach has also been evaluated using certain performance metrics, including Accuracy, Precision, Recall, and F1-score. These performance metrics give an overall idea of how well the classification approach is performing, especially for tasks such as medical image classification, where both false positives and false negatives play an important role. The results obtained using the proposed approach for each of the five folds of the dataset are presented in TABLE 2. From TABLE 2, it is clear that the proposed approach has been performing extremely well for all folds.

Table 2: Five-Fold Cross-Validation Result of Nodule-Net

Fold	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Fold 1	99.96	99.95	99.96	99.95
Fold 2	99.97	99.97	99.97	99.97
Fold 3	**99.98**	**99.98**	**99.98**	**99.98**
Fold 4	99.97	99.96	99.97	99.96
Fold 5	99.96	99.95	99.96	99.95

The results of the cross-validation process reveal that the Nodule-Net model performs with consistent high performance and low variance in performance across different folds, thereby ensuring the robustness and overfitting resistance of the model. The minor variations in performance across different folds ensure the stability of the learning process, and the performance of 99.98% on the best fold in all metrics again validates the efficiency of the proposed hybrid architecture in extracting discriminative features for accurate classification.

3.4 Comparative Analysis

To validate the effectiveness of the proposed Nodule-Net framework, a detailed comparative analysis has been carried out with various state-of-the-art deep learning models presented in TABLE 3. The comparative analysis has been carried out on the basis of various evaluation metrics such as Accuracy, Precision, Recall, and F1-score, ensuring a fair assessment of the models. It is clear that the proposed Nodule-Net framework has achieved better results compared to the existing models in terms of all the evaluation metrics. It has been observed that traditional models such as VGG-16 and ResNet-50 achieve higher recall, but the precision and accuracy are relatively low, indicating the limitations of these models in balancing the false positive and false negative rates. More advanced models such as Xception and ensemble models achieve relatively better results, but these models are not able to achieve optimal consistency in all the metrics.

Notably, models incorporating temporal components, such as VGG-19 + LSTM, demonstrate strong performance (up to 99.42% accuracy), highlighting the importance of sequential learning. Neverthe-

Table 3: Comparative Performance of Proposed Methods with Existing Methods for Chest CT Dataset

Method	**Accuracy %**	**Precision %**	**Recall %**	**F1 %**
VGG-16 [27]	84.40	83.58	99.68	90.49
Xception [27]	95.80	96.30	99.50	97.39
ResNet 50 [27]	78.80	78.92	99.62	88.04
Custom CNN [28]	93.06	95.53	93.09	93.84
VER-Net [29]	91.00	91.00	91.00	91.00
VGG 19+LSTM [30]	99.42	99.18	99.16	99.12
ResNet101+Inception V3 [31]	93.70	83.85	82.27	82.45
CNN + Inception V3, Xception, and ResNet-50 [32]	92	92	91.72	91.74
Proposed	**99.98**	**99.98**	**99.98**	**99.98**

less, there is a lack of overall global spatial dependency and sophisticated preprocessing techniques. The main reason behind the performance of Nodule-Net is based on the hybrid model that uses CNN and transformer architectures. These two architectures are used in tandem to effectively capture local and global spatial dependency. In addition, there is a sophisticated preprocessing technique that ensures the input data is clean, standardized, and of high quality. Moreover, there is an incorporation of temporal modeling that allows the analysis of the progression of nodules. This synergy of components leads to the development of a robust, accurate, and well-generalized framework in the classification of lung cancer. It can therefore be concluded that the proposed Nodule-Net achieves state-of-the-art performance by having the best accuracy, precision, recall, and F1-score compared to all other techniques. It can therefore be concluded that the proposed hybrid technique is more effective in the classification of lung cancer compared to existing techniques.

FIGURE 10 shows the comparative results of all models for the Chest CT dataset. From the figure, it is clear that there is an upward trend as we move from traditional CNNs to hybrid and transformer architectures. Although traditional models have shown moderate accuracy and F1 scores, more recent transformer architectures have shown results close to 95-98%. The proposed method has been shown at the top of all models, indicating perfect results for all four parameters. This shows that the proposed method is highly efficient in accurately classifying chest CT images and has shown good results for all diverse images.

3.5 Ablation Study

In order to effectively quantify the contribution of each architectural component of the Nodule-Net framework, the ablation study was conducted by progressively enabling the key components of the network, such as the preprocessing step, local feature extraction, global context modeling, and the temporal analysis.

The ablation study configurations are as follows:

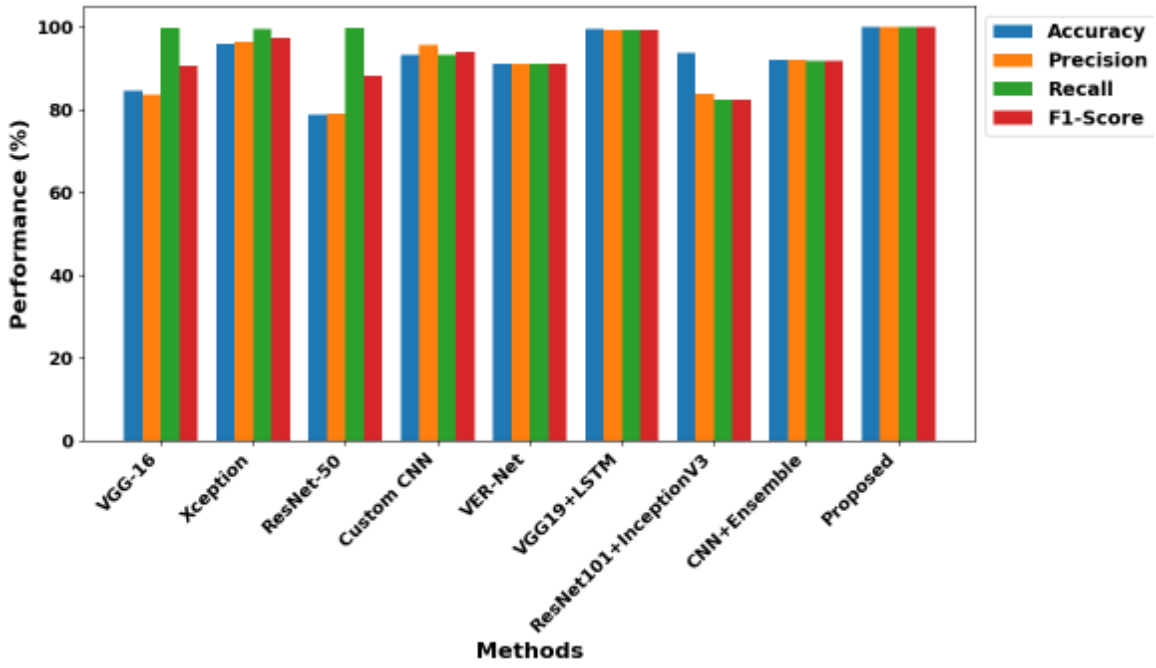


Figure 10: Comparative Performance Analysis

- Baseline (CNN Only): This configuration includes the basic CNN network.
- + Preprocessing: This configuration includes the resampling step, intensity normalization, CLAHE, Non-Local Means filter, and brightness adjustment.
- + Local Feature Module (NoduleVisionNet): This configuration includes the residual-based hierarchical feature extraction.
- + Global Context Module: This configuration includes the transformer-based long-range dependency modeling.
- + Temporal Modeling: This configuration includes the longitudinal analysis of the progression of the nodule.
- Full Model (Nodule-Net): This configuration includes the complete hybrid network.

This is clear from the result obtained in TABLE 4, which indicates that all components of the proposed Nodule-Net framework contribute progressively towards the improvement in performance. Starting from the baseline CNN, the addition of advanced preprocessing techniques improves the quality of the input, followed by significant improvements in the results after the addition of the local feature extraction module. Then, the addition of global context modeling further improves the classification results, while the addition of temporal modeling refines the results by incorporating patterns of nodule progression. The results for the proposed model are 99.98%, as expected, since all metrics are integrated into the proposed hybrid framework, which has been shown to have a synergistic effect.

Table 4: Ablation study of Nodule-Net

Configuration	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Baseline (CNN only)	96.42	96.10	96.35	96.22
+ Preprocessing	97.85	97.60	97.80	97.70
+ Local Feature Module	98.96	98.75	98.90	98.82
+ Global Context Module	99.52	99.40	99.50	99.45
+ Temporal Modeling	99.78	99.70	99.76	99.73
Full Model (Nodule-Net)	**99.98**	**99.98**	**99.98**	**99.98**

3.6 Discussion

This work presents a comprehensive evaluation of a proposed hybrid DL architecture for lung cancer detection and classification, as validated on the Chest CT datasets. The proposed DL architecture comprises an efficient preprocessing scheme, local feature enhancement, global context information, and combined feature integration for effective lung cancer detection and classification. The preprocessing scheme includes standardizing image resolutions, intensity normalization, contrast enhancement using CLAHE, Non-Local Means filtering, and brightness enhancement. This ensures the quality of input images, minimizes noisy data, and enhances the reliability of the proposed DL framework. The proposed DL framework was compared with existing popular approaches, demonstrating its efficiency in lung cancer detection and classification. The prevailing approaches, such as VGG-16, Xception, ResNet50, Custom CNN, VGG19-LSTM, CNN-InceptionV3, achieved fair results. However, the proposed DL framework surpassed all the prevailing approaches. On the Chest CT dataset, the proposed DL framework achieved 99.98% accuracy, precision, recall, and F1-score, which is greater than all the prevailing approaches. These results highlight the ability to differentiate benign and cancerous nodules effectively and demonstrate the generalization ability of the framework on datasets with varying image features. The ablation study also justifies the contribution of individual components to the overall framework. The baseline showed a mediocre performance, but this has been significantly improved with the addition of the pre-processing pipeline. The enhancement in local feature extraction has allowed small and irregular nodules to be easily identified, while the global features were helpful in identifying patterns that indicate the existence of multiple nodules. The hybridization of local and global features has shown the most significant improvement in the overall performance, thereby highlighting the importance of local features that provide fine-grained information as well as global features that provide background information. The proposed framework has demonstrated state-of-the-art results with the best accuracy, robustness, and interpretability. The feature fusion, pre-processing, and interpretability aspects have highlighted a robust framework that is generalizable and applicable to early lung cancer screening.

4. CONCLUSION

The proposed Nodule-Net framework offers a promising and high-performing framework for the detection and classification of lung cancer by effectively combining local feature learning, global context learning, and the dynamics of nodule evolution over time. The framework demonstrates

near-perfect prediction performance with the advantage of computational efficiency, which can positively impact the diagnosis of lung cancer and the development of automated medical image analysis techniques in real-world healthcare environments. However, the proposed framework was evaluated using limited datasets, which may limit its generalization capacity across various real-world environments due to the diversity of clinical conditions and datasets. Future work aims to validate the proposed framework using large-scale datasets and heterogeneous datasets with the help of domain adaptation techniques to improve the generalization capacity of the proposed framework. Furthermore, the integration of Explainable AI techniques can also improve the usability of the proposed framework in real-world environments.

5. FUNDING

No funding was received for this work.

6. AUTHORS' CONTRIBUTIONS

All authors reviewed the results and approved the final version of the manuscript.

References

[1] Salama WM, Shokry A, Aly MH. A Generalized Framework for Lung Cancer Classification Based on Deep Generative Models. Multim Tools Appl. Springer Nature. 2022;81:32705-32722.

[2] Kasinathan G, Jayakumar S, Gandomi AH, Ramachandran M, Fong SJ, et al. Automated 3-D Lung Tumor Detection and Classification by an Active Contour Model and CNN Classifier. Expert Syst Appl. 2019;134:112-119.

[3] Miller KD, Nogueira L, Devasia T, Mariotto AB, Yabroff KR, et al. Cancer Treatment and Survivorship Statistics, 2022. CA Cancer J Clin. 2022;72:409-436.

[4] Guo Y, Song Q, Jiang M, Guo Y, Xu P, Zhang Y et al. Histological Subtypes: Classification of Lung Cancers on CT Images Using 3D Deep Learning and Radiomics. Acad Radiol. 2021;28:e258-e266.

[5] Gunasekaran KP. Leveraging Object Detection for the Identification of Lung Cancer. 2023. ArXiv preprint: https://arxiv.org/pdf/2305.15813.

[6] Zhang G, Yang Z, Gong L, Jiang S, Wang L, et al. Classification of Lung Nodules Based on CT Images Using Squeeze-And-Excitation Network and Aggregated Residual Transformations. La Radiologia Medica. 2020;125:374-383.

[7] Jena SR, George ST, Ponraj DN. Lung Cancer Detection and Classification With DGMM-RBCNN Technique. Neural Comput Appl. Springer Nature. 2021;33:15601-15617.

[8] Pandian R, Vedanarayanan V, Ravi Kumar DN, Rajakumar R. Detection and Classification of Lung Cancer Using CNN and Google Net. Meas Sens. 2022;24:100588.

[9] Zhang C, Sun X, Dang K, Li K, Guo XW, et al. Toward an Expert Level of Lung Cancer Detection and Classification Using a Deep Convolutional Neural Network. The Oncologist. 2019;24:1159-1165.

[10] Kanavati F, Toyokawa G, Momosaki S, Rambeau M, Kozuma Y, et al. Weakly-Supervised Learning for Lung Carcinoma Classification Using Deep Learning. Sci Rep. 2020;10:9297.

[11] Bicakci M, Ayyildiz O, Aydin Z, Basturk A, Karacavus S, et al. Metabolic Imaging Based Sub-Classification of Lung Cancer. IEEE Access. 2020;8:218470–218476.

[12] Shin H, Oh S, Hong S, Kang M, Kang D, et al. Early-Stage Lung Cancer Diagnosis by Deep Learning-Based Spectroscopic Analysis of Circulating Exosomes. ACS Nano. 2020;14:5435-5444.

[13] Hussain Ali YH, Chinnaperumal S, Marappan R, Raju SK, Sadiq AT, et al. Multi-Layered Non-Local Bayes Model for Lung Cancer Early Diagnosis Prediction With the Internet of Medical Things. Bioengineering. 2023;10:138.

[14] Li T, Lin Q, Guo Y, Zhao S, Zeng X, et al. Automated Detection of Skeletal Metastasis of Lung Cancer With Bone Scans Using Convolutional Nuclear Network. Phys Med Biol. 2022;67:015004.

[15] Feng J, Jiang J. Deep Learning-Based Chest CT Image Features in Diagnosis of Lung Cancer. Comput Math Methods Med. 2022;2022:1-7.

[16] Shimazaki A, Ueda D, Choppin A, Yamamoto A, Honjo T, et al. Deep Learning-Based Algorithm for Lung Cancer Detection on Chest Radiographs Using the Segmentation Method. Sci Rep. 2022;12:727.

[17] SU A, PP FR, Abraham A, Stephen D. Deep Learning-Based BOVW–CRNN Model for Lung Tumor Detection in Nano-Segmented CT Images. Electronics. 2023;12:14.

[18] Kasinathan G, Jayakumar S. Cloud-Based Lung Tumor Detection and Stage Classification Using Deep Learning Techniques. Biomed Res Int. 2022;2022:4185835.

[19] Ragab M, Abdushkour HA, Nahhas AF, Aljedaibi WH. Deer Hunting Optimization With Deep Learning Model for Lung Cancer Classification. Comput Mater Contin. 2022;73:533-546.

[20] Wang X, Wang L, Zheng P. Sc-Dynamic R-CNN: A Self-Calibrated Dynamic R-CNN Model for Lung Cancer Lesion Detection. Comput Math Methods Med. 2022;2022:9452157.

[21] Ren Z, Zhang Y, Wang S. A Hybrid Framework for Lung Cancer Classification. Electronics. 2022;11:1614.

[22] Wankhade S, S. A Novel Hybrid Deep Learning Method for Early Detection of Lung Cancer Using Neural Networks. Healthcare Analytics. 2023;3:100195.

[23] Cao Y, Liu L, Chen X, Man Z, Lin Q, et al. Segmentation of Lung Cancer-Caused Metastatic Lesions in Bone Scan Images Using Self-Defined Model With Deep Supervision. Biomed Signal Process Control. 2023;79:104068.

[24] Tashtoush Y, Obeidat R, Al-Shorman A, Darwish O, Al-Ramahi M, et al. Enhanced Convolutional Neural Network for Non-Small Cell Lung Cancer Classification. Int J Power Electron Drive Syst. 2023;13:1024-1038.

[25] Zhang Y. Lung Segmentation With NASNet-Large-Decoder Net. 2023. ArXiv Preprint: https://arxiv.org/pdf/2303.10315

[26] https://www.kaggle.com/datasets/mohamedhanyyy/chest-ctscan-images

[27] Golkarieh A, Kiashemshaki K, Boroujeni SR, Isakan NA. Advanced U-Net Architectures With CNN Backbones for Automated Lung Cancer Detection. 2025. ArXiv Preprint: https://arxiv.org/pdf/2507.09898

[28] Hammad M, ElAffendi M, El-Latif AAA, Ateya AA, Ali G, et al. Explainable AI for Lung Cancer Detection via a Custom CNN on CT Images. Sci Rep. 2025;15:12707.

[29] Saha A, Ganie SM, Pramanik PK, Yadav RK, Mallik S, et al. VER-NET: A Hybrid Transfer Learning Model for Lung Cancer Detection Using CT Scan Images. BMC Med Imaging. 2024;24:120.

[30] Alsheikhy AA, Said Y, Shawly T, Alzahrani AK, Lahza H. A CAD System for Lung Cancer Detection Using Hybrid Deep Learning Techniques. Diagnostics. 2023;13:1174.

[31] Akhand ZEN, Rahman AI, Sarda A, Fahim MZA, Tushi LT, et al. Lung Cancer Detection Using Ensemble Technique of CNN. In Proceedings of International Conference on Information and Communication Technology for Development: ICICTD. Singapore: Springer. 2023:497–507.

[32] Mamun M, Mahmud MI, Meherin M, Abdelgawad A. LCDctCNN: Lung Cancer Diagnosis of CT Scan Images Using CNN-Based Model. In 2023 10th International Conference on Signal Processing and Integrated Networks. SPIN. IEEE. 2023:205–212.