Scientific Reports

https://doi.org/10.1038/s41598-026-41161-7

Article in Press

A hybrid deep learning approach integrating CNN and transformer for lung cancer classification using CT scans

Received: 30 July 2025

Accepted: 18 February 2026

Published online: 17 March 2026

Cite this article as: Yousafzai S.N., Nasir I.M., Mansour S. *et al.* A hybrid deep learning approach integrating CNN and transformer for lung cancer classification using CT scans. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-41161-7

Samia Nawaz Yousafzai, Inzamam Mashood Nasir, Sahar Mansour, Noha Negm, Asma A. Alhashmi, Mohannad A. Alharbi & Eunchan Kim

A Hybrid Deep Learning Approach Integrating CNN and Transformer for Lung Cancer Classification Using CT Scans

Samia Nawaz Yousafzai[1,*], Inzamam Mashood Nasir[2], Sahar Mansour[3], Noha Negm[4], Asma A. Alhashmi[5], Mohannad A Alharbi[6], and Eunchan Kim[7,*]

[1]Department of Computer Science, HITEC University Taxila, Pakistan; samia.nawaz@hitecuni.edu.pk
[2]Faculty of Informatics, Kaunas University of Technology, 51368 Kaunas, Lithuania; inzamam.nasir@ktu.edu
[3]Department of Radiological Sciences, College of Health and Rehabilitation Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia; sabdelety@pnu.edu.sa
[4]Department of Computer Science, College of Science & Art at Mahayil, King Khalid University, Saudi Arabia; nohaabdulhamid@kku.edu.sa
[5]Department of Computer Science, College of Science, Northern Border University, Arar, Saudi Arabia; asalhashmi@nbu.edu.sa
[6]Department of Information Science, College of Humanities and Social Sciences, King Saud University, P. O Box 28095, Riyadh 11437, Saudi Arabia; mhanalharbi@ksu.edu.sa
[7]Department of Information Systems, Hanyang University, Seoul, Republic of Korea; eckim@hanyang.ac.kr
[*]Corresponding authors: samia.nawaz@hitecuni.edu.pk (S.N.Y.) and eckim@hanyang.ac.kr (E.K.)

ABSTRACT

Lung cancer is an extremely fatal kind of cancer, resulting in the deaths of almost 7.6 million individuals annually around the globe. Nevertheless, a timely diagnosis is a crucial necessity for enhancing the likelihood of human survival. Regarding tumor identification, CT scans are normally used to identify affected areas. Nevertheless, CT imaging face significant problems such as poor visibility of tumor locations and high false negative rates. The small dataset size of medical imaging makes it challenging to capture local lesion features by iterative training, considering all input features equally. This work integrates Convolutional Neural Network (CNN) and Improved Swin Transformer (C-Swin), a deep learning model that extracts and integrates fine-grained local and global features. C-Swin has Transformer encoder and a CNN module. The CNN module extracts local features, whereas the Transformer module captures global features. The Transformer encoder uses a hybrid shifted window attention method to focus on a spatial region of the CT image, reducing background semantic information and improving local feature capture accuracy. The proposed method is validated using the publicly accessible Kaggle dataset namely IQ-OTH/NCCD with three classes. the proposed C-Swin model achieved average accuracy of 96.26%, precision of 97.48%, recall of 96.39% and f1-score of 97.42%. The numerical findings unequivocally demonstrate that our proposed method surpasses various existing methods with an increase in accuracy ranging from 2.31% to 6.81%. The C-Swin model is capable of extracting detailed local lesion features, resulting in improved classification performance.

Introduction

Lung cancer remains one of the most severe and life-threatening diseases worldwide[1,2]. According to recent reports by the World Health Organization (WHO), it is responsible for nearly 7.6 million deaths annually across the globe[3,4]. Moreover, the incidence of cancer is anticipated to grow, reaching an estimated 17 million individuals worldwide by 2030[5]. Lung cancer is diagnosed primarily after the age of 50; consequently, the proportion of lung cancer patients is steadily increasing. The sole method of curing is to identify it in its initial stages[6]. Medical imaging possesses remarkable capabilities for the cure of clinical and noninvasive analyses in the domain of Computer Vision (CV)[7,8]. The primary application of the obtained imaging modalities are MRI, X-rays, and CT use to diagnose a specific disease[9,10]. The CT is a renowned medical modality that records images on film[11,12]. The radio waves are employed to observe the contrast of MRI images, while the X-rays are employed to observe the CT scan. The majority of physicians opt for CT scans for their patients because of the extreme computational expense of MRI in comparison to CT scan. The primary benefit of a CT scan is its ability to quickly capture the thin structure, tissues, and cerebral organs.

Lung cancer is extremely fatal diseases due to the difficulties related with its detection compared to other diseases. The

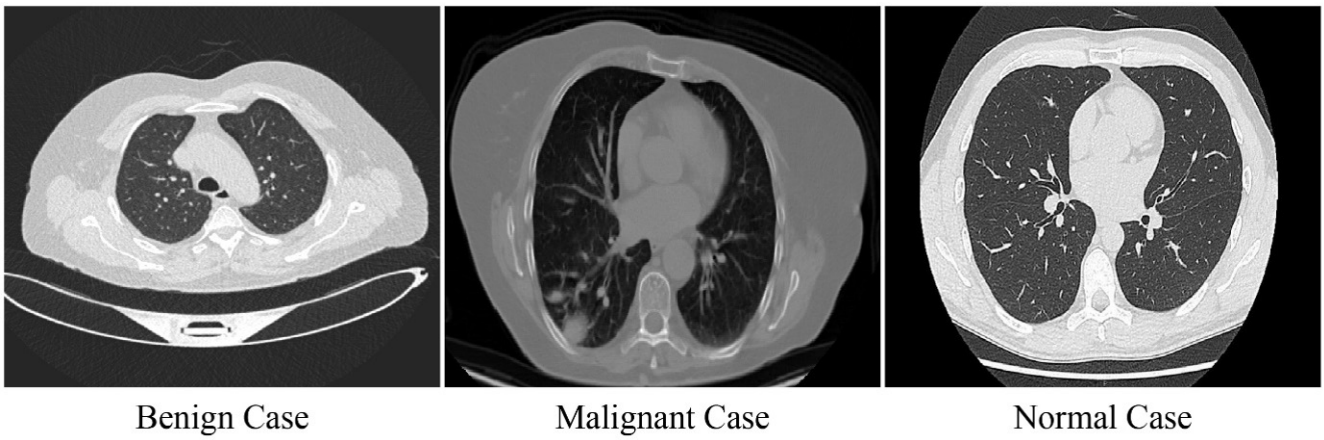

Figure 1. Sample images from selected dataset

primary cause of failure is the diminutive lesion size, which is sometimes referred to as a nodule. At first stage, the size of cancer cells is very little. However, with time, the tumor size progressively increases, leading to its transformation into a malignant state. The radiologist accurately identified the condition at that time; however, it was unfortunately too late for the individuals. It has become imperative to manage the disease at its early stage. The likelihood of survival rate can be intensified if the condition is detected in its initial stages[13]. Recently, scientists in the field of CV have developed automated algorithms that can recognize and categorize both malignant and healthy regions[14–16].

Recently, Deep Learning (DL) and Machine Learning (ML) have been implemented in the field of medical imaging[17–19], precision agriculture[20–22] and language processing[23–26]. A variety of contemporary conventional methods are being implemented to classify medical images[27–29]. These techniques adhere to a four-step structure, encompassing preprocessing to classification[30–32]. Preprocessing is crucial in medical imaging when using traditional approaches, as the original CT scans may are noisy that produce negative characteristics[16,33–35]. The presence of negative traits hampers the accuracy of categorization, making the elimination of irrelevant features a significant task. The pulmonary nodule is of minute dimensions, making it challenging to distinguish between the normal, benign and malignant images using feature-based approaches. Figure 1 displays sample CT scans of benign, malignant, and normal condition.

The primary obstacles encountered by the researcher in this study are the resemblance among normal and cancerous conditions. CT scans produce similar features that impact performance of classification. The appearance of lung nodules in the images is minimal, making it difficult to distinguish between healthy and cancerous images. Selecting and reducing the most relevant features is one of the major challenges since using inappropriate features increases the error rate significantly. Therefore, the aim is to minimise the total error rate.

The proposed C-Swin is a DL model that has been specifically developed for the classification of lung cancer. The C-Swin model is intended to enable efficient extraction and integration of local fine-grained information about lung CT scans. C-Swin combines the strengths of CNNs for detailed low-level feature extraction with the strengths of the Transformer model to obtain global contextual features. The transformer encoder uses an attention mechanism that modifies the position of the shifted window to focus the model on the area of interest, which in this case is the lungs, without introducing extraneous background semantic data. As a result of the attention mechanism, C-Swin maintains the model's ability to accurately identify local features in the lung CT scan image versus irrelevant information. The second part of the attention window mechanism preserves the boundary relationships between the adjacent local feature regions. The C-Swin model, therefore, increases the efficiency of the extraction of local spatial features from CT datasets, thereby creating a stronger relationship between local lesions and lung cancer classification. The primary contributions of this article are outlined as follows:

- A DL model that integrates CNN and transformer is proposed to accurately capture local and global features of CT images more effectively in lung cancer patients. The "IQ-OTH/NCCD" dataset is used to train and evaluate the C-Swin model.
- To optimize the extraction of CT-localized features, a hybrid shifted window attention mechanism is utilized in the Swin Transformer.
- The model's efficacy was evaluated in the context of various data enhancement techniques. In the interim, the model's features extracted in classification tasks were analyzed and visualized. The results indicated that the model could

concentrate on lesion regions that are closely associated with lung cancer, a capability that is particularly valuable in clinical diagnosis.

In order to facilitate the reader's comprehension of the investigation, the paper is divided into numerous sections. Section 2 offers an overview of pertinent research. The framework and methodologies employed in this research are elucidated in Section 3, while the experimental outcomes are reported and analysed in Section 4. Section 5 provides a concise interpretive discussion of the results. The manuscript is concluded by Section 6, which emphasises the study's limitations and suggests potential areas for future research.

Related Work

Globally, lungs cancer ranks among the main causes of human mortality[36]. Several recent studies on the basis of lung cancer diagnosis using DL are discussed here. A CNN incorporating multi scale framework was introduced with transfer learning technique to classify pulmonary nodules in CT images into two categories[37]. Another study employed supervised learning multi scale framework to efficiently capture both intrascale and interscale contextual information of pulmonary lesion[38]. Pretrained models are employed to capture features of pulmonary lesion. They effectively classified pulmonary nodules as benign or malignant by integrating a bag of words model with various 2D slice features[39]. An attribute based Generative Adversarial Network (GAN) was developed that enhances categorization and generation performance through the use of self-attention processes and numerous loss functions[40]. DL framework based on convolutional self-encoder was proposed, which employed classifier based on clustering to classify lung nodules. They employed GAN as the data enhancement method[41].

Recent development in medical image classification have led to the development of various DL models across different domains of healthcare. For example, the CICADA (UCX) model for breast cancer segmentation introduces a novel approach to enhancing diagnostic precision by leveraging ML techniques[42]. Similarly, a hybrid model combining ML and DL, has shown impressive performance in predicting anti-cancer drug responses, suggesting potential applications for optimizing treatment prediction in lung cancer therapy[43]. The TATHA model has been the best performing model for thyroid nodule detection using ultrasound images[44]. The AKAttNet framework provides a detailed description of the performance of various feature selection techniques in the classification of autism spectrum disorder[45]. The PABT-Net also provides a novel framework for classifying brain tumors that uses hierarchical attention mechanisms for fine-grained classification and supports the need to incorporate global and local features to categorize lung cancer accurately[46]. Models from DNN's for diabetic retinopathy detection have also demonstrated the importance of accurately extracting features from medical images in order to classify[47]. The UIGO model, used for liver tumor segmentation, has demonstrated high precision and efficiency while also confirming the role of hybrid DL approaches when using automated methods for medical image segmentation[48]. The BCB-CSPA network demonstrates the efficiency of optimised DL systems for efficient classification of skin cancer[49]. Moreover, the MRA-Net, a attention based multiscale-CNN for Alzheimer's Disease classification, offering a promising attention-based approach for improving the early detection of lung cancer[50]. The PAM-UNet Framework combines the Parallel Attention Module to increase the quality of both extraction of features from medical images and classifying them accurately[51].

LungNet, a shallow CNN, was developed which comprises of two CNNs models. The weights are shared through transfer learning. Nevertheless, the procedure extracts features using a single-scale filter. Although, the use of multi-scale filters to analyze nodule images could produce more effective nodule features, as a result of the variation in nodule diameters. In comparison to conventional feature extraction methods, it exhibits superior accuracy[52]. The Swin Transformer model's performance was examined for both lung cancer classification and segmentation tasks. The pre-trained Swin-B model achieved a top-1 classification accuracy of 82.26%, exceeding the Vision Transformer (ViT) by 2.529%. The Swin-S model surpassed previous models in segmentation as measured by mean Intersection over Union (mIoU)[53]. A lung cancer detection system that utilizes CNN and NLP. Deep feature-based CNN were specifically utilized to classify lung cancer tumors, achieving an accuracy rate of 88%. Additionally, the system was enhanced with a chatbot that utilizes NLP, methods to offer instant details. It offers a superior accuracy and a lower level of precision[54]. A two-branch model that combines CNN with Data-efficient Vision Transformer, for the purpose of classifying non-small cell lung cancer using CT Images. The model achieved high performance with 96.10% accuracy, on publicly available data from Kaggle[55].

A Computer Aided lung cancer detection technique was proposed using CNNs to categorize CT images. The paper focused on how the CAD system works through a DL model and that CNNs can detect malignant lung cancers in CT images by using two classes of malignant and non-malignant. Segmentation and preprocessing techniques were employed to accomplish classification. It has a high value for the area under the curve and a low value for precision[56]. Lung cancer detection was performed using transfer learning with the GoogLeNet model. The given approach was examined using the IQ-OTH/NCCD dataset. A Deep Neural Network (DNN) was employed to identify malignant lung nodules in CT images. The model applies quick preprocessing to input images, which involves separating the region of interest, which is primarily composed of lung tissue, and removing images of adjacent tissues and artifacts. It offers a high value for the area under the curve and a low value

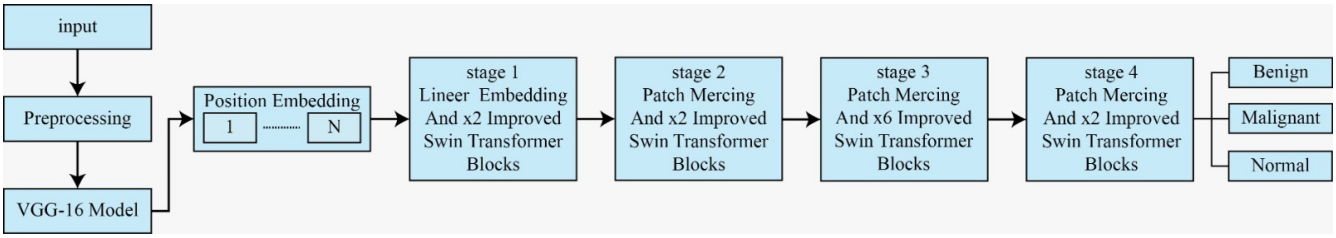

Figure 2. Architecture of proposed C-Swin model

for precision[57]. Another study utilized IQ-OTH/NCCD dataset to correctly detect lung cancer based on CT scans. The data was collected from Iraqi hospitals and classified into three classes using a CNN that included the AlexNet architecture. The dataset provided high precision but low accuracy[58]. The DY-FSPAN model for detection of lung cancer using histopathological images, employing the Cheetah Hunting Algorithm (CHA) for feature selection as well as a dilated Y-Block structure and pyramid attention mechanism to enhance the classification of tissue samples. The DY-FSPAN model showed 98.5% accuracy based on the LC25000 dataset of lung and colon cancer as well as associated tissue characteristics[59].

Due to their visual nature, analyzing medical images is a challenging and intriguing endeavor. The literature presents various ways for addressing the spread of powerful tissues and tumors, but their effectiveness is limited. Lung cancer is a cell disease that is caused by the irregular development and growth of cells, which leads to the formation of a tumor. Lymph fluid envelops the lungs or pulmonary tissue, serving as a barrier against the infiltration of cancer cells into the lungs. DL is more prevalent than conventional methods in the field of image classification. The mathematical approach has facilitated clinical decision-making for lung cancer. The proposed technique accurately identifies the expected lung cancer cells.

Proposed Methodology

This section describes the proposed method designed to classify lung cancer based on CT imaging data. The entire sequence of steps in the suggested methodology is illustrated in Figure 2. The CT scan images are first retrieved from the dataset. Subsequently, a series of pre-processing procedures are carried out on the images. Obtaining a significant amount of annotated data for training the model is difficult because of the characteristics of medical imaging. The quantity of training samples is artificially increased through the use of data augmentation. The proposed methodology employs a convolutional transformer-based approach to classify lung cancer.

Data Preprocessing

Before the training and testing, the dataset undergoes preprocessing operations. Initially, photos from each category are randomly shuffled to ensure equitable learning. The photos are divided into a distribution of 70% for training, 20% for testing, and 10% for validation. Thus, 70% of the complete collection of images is selected randomly to constitute the training set, utilised for model training. 20% of the photographs are randomly chosen and allocated to the testing set, while 10% are designated for validation. The testing set is employed to assess the model's efficacy on new, unobserved images. The initial CT image was altered to eliminate extraneous aspects, like background and noise, to ensure precise training. Space normalisation and bias field correction are employed to enhance photograph quality by eliminating undesirable fluctuations and distortions. The preprocessed photos undergo data augmentation. Data augmentation methods such as horizontal flipping, zooming, rotating, scaling, and shearing are employed to enhance the dataset by augmenting the sample size. The cumulative sum of augmented CT scans is 3598. Of the total, 968 instances are categorised as benign, 1400 cases as malignant, and 1230 cases as normal.

CNN Architecture

The C-Swin model utilizes each CT scan as input, with each image measuring $256 \times 256 \times 3$. The CNN module is based on VGG16[60] which comprises of 13 convolution layers, as illustrated in Figure 3. Bilinear interpolation algorithm is used to scale each image to $112 \times 112 \times 3$ align with the input size of feature maps of VGG16. Variety of filters with varying sizes are present in convolutional layers to capture necessary features such as texture and shape from CT scans. The output feature representation is $3 \times 3 \times 512$ after 5 pooling, 13 convolutions and activations operations in VGG16 are performed. Lastly, a convolutional layer is incorporated at the ending of VGG16 to facilitate the feature mapping to represent features. A 512 in-channel and 256 out-channel are set up for this convolutional layer. Tokens of 256 dimensions, obtained after the last convolutional layer, represent the CT sample. Despite the fact that more powerful CNN networks such as EfficientNet and DenseNet may optimize parameters more effectively, VGG16 was chosen for its stability and good performance rates on tiny medical datasets. Its unified and simple architecture allows for easy integration with the Transformer module in our hybrid C-Swin framework.

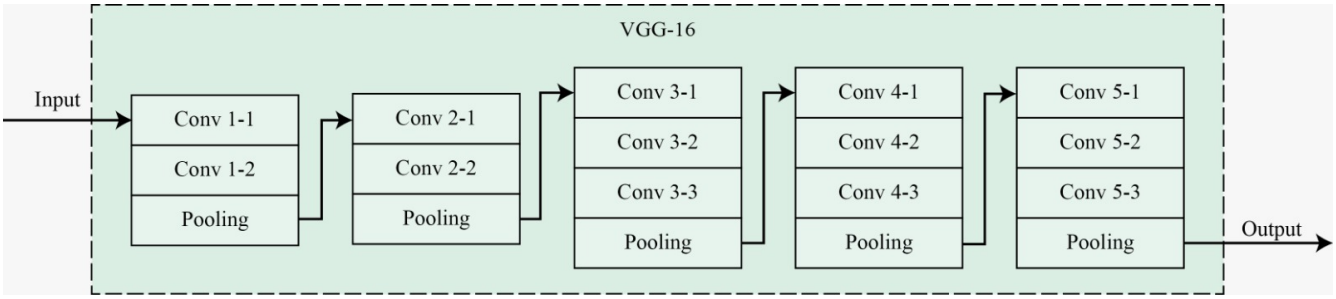

Figure 3. Architecture of VGG16 model

Given the small size of the IQ-OTH/NCCD dataset, more complex or profound architectures are likely to be overfitted. VGG16 trades off depth and generality to make it suitable for robust local feature extraction. Furthermore, the model is unique in that it combines CNN and hybrid self-attention mechanisms, with VGG16 serving as an appropriate supplement to Transformer's global context learning.

Transformer Architecture

The transformer's structure is based on an encoder and a decoder, and it was originally developed for language translation. The primary concept of transformer is that the Self Attention Mechanism (SAM)[61] is the sole mechanism that can be employed to infer dependencies between input and output. The input sentence is incorporated into a token for the purpose of language translation. This input token sequence is presented to the encoder, which transforms it into a vector of fixed size. Subsequently, the decoder converts this vector into an output sequence. The encoder serves as a weighted fusion mechanism in Transformer-based image classification tasks. Each slice is depicted as a visual token, and the image is segmented. These visual tokens are transmitted to the encoder for mutual feature fusion, which generates an output that is immediately transmitted to the classification head for classification. In contrast to translation tasks, image classification tasks do not necessitate a decoder for the transformer. Consequently, the encoder in transformer is applied exclusively in this article.

Self-Attention Mechanism

The Transformer's fundamental component is the SAM, which is helpful for the weighted fusion of all input image patches. Initially, a linear transformation is applied to every patch in the input sequence to generate a collection of Query (Q), Key (K), and Value (V). For each patch sequence, the Q vector is employed to multiply with the K of other patch sequence to determine the similarity score of this patch sequence with the other patches in the sequence. Next, similarity scores are multiplied with the V of other sequence to derive the weighted fusion of the patches with all patches in the sequence. The formula for SAM is as follows:

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_K}}\right)V \tag{1}$$

where the dimension of embedding K is denoted as d_K. As the dimensionality of the vector K rises, the size of the dot product will likewise increase, resulting in a minimal gradient and a tendency for the softmax activation function to approach its saturation region. Hence, the output of the dot product between vectors K and Q is divided by $\sqrt{d_K}$. To avoid the disappearance of the gradient.

S-GLU Based MLP

S-GLU integrates the Swish (S) activation function inside its architecture, hence facilitating remarkable progress in neural network design, exemplified by the swin transformer[62]. The invention relies solely on the distinction between the gated mechanism and input processing, setting it apart from conventional GLU designs. This separation confers a benefit to S-GLU: enhanced information flow throughout the network and targeted modification of feature representations. The outcomes of employing the swin transformer design, in conjunction with S-GLU, for lung cancer detection have been exceptional. In contrast to alternatives such as ReLU, S-GLU employs the S activation function, which is distinguished by its nonlinear characteristics and smoother gradients. S-GLU improves the model's capability to identify nuanced patterns in lung cancer images compared to other models, further augmented by its gating mechanism that dynamically amplifies or suppresses features at various levels of hierarchical processing, thereby refining the differentiation between malignant and benign lesions.

$$z_g = W_g \times x + b_g \tag{2}$$

$$z_x = W_x \times x + b_x \tag{3}$$

$$S(z_x) = z_x \times \sigma(z_x) \tag{4}$$

$$y = S(z_x) \odot z_g \tag{5}$$

$$o = W_2 \times y + b_2 \tag{6}$$

The precision capacity of S-GLU is enhanced by the unique combination of feature qualities from the (S) activation function and the selective gating mechanism of GLU. The collaboration between the two components is likely to improve the Swin Transformer's capability to discern intricate details in photos of lung cancer, potentially enhancing its power to identify tiny indications of cancer. Consequently, S-GLU may serve as a valuable adjunct to DL models in medical image processing. As seen in Figure 4, the S-GLU-based MLP architecture facilitates more efficient training and superior generalization skills.

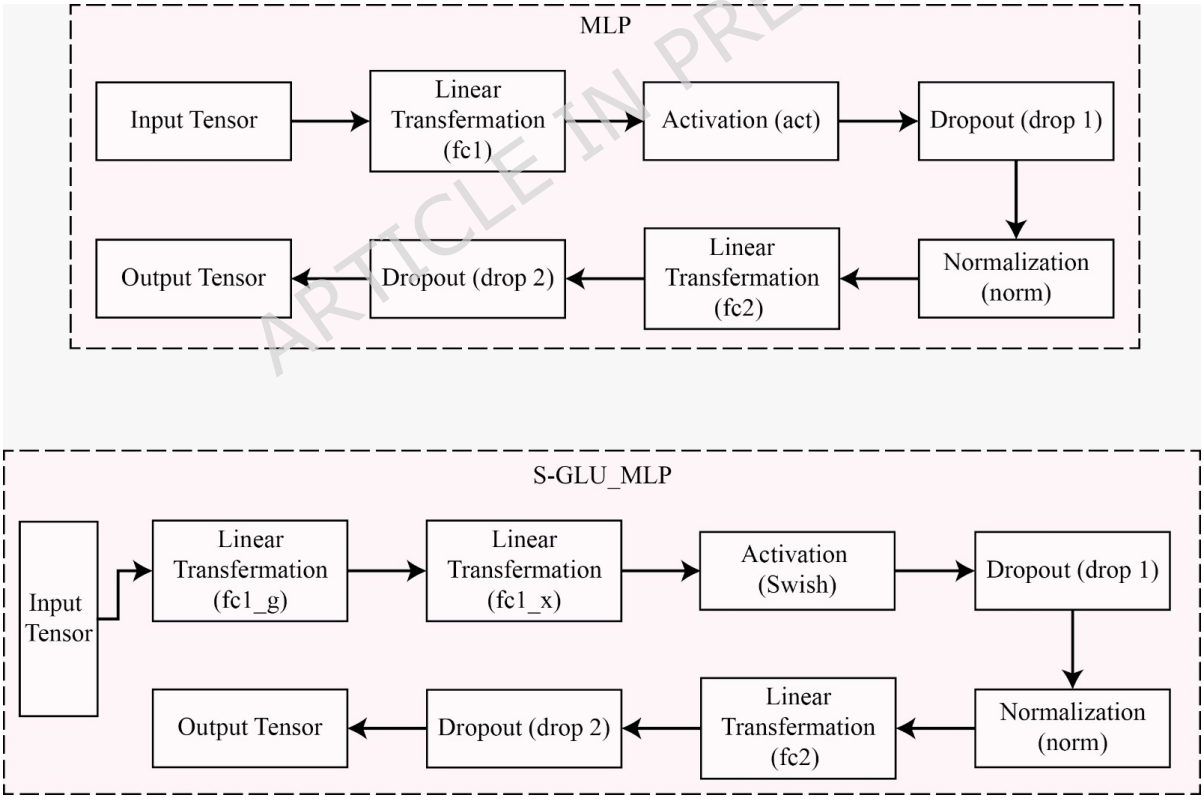

Figure 4. Architecture of general MLP and S-GLU based MLP

Hybrid Shifted Window Based Multi Head Self Attention (HSWMHA)

Two multi-head self-attention methods, window MHA and shifting window MHA, are incorporated into Swin models. Using a hybrid shifting window method, the Swin-Base model makes use of hybrid Swin Transformer blocks. HSWMHA is a

version of standard transformer self-attention that integrates window-based and shifting techniques, significantly enhancing the efficiency of processing large-scale data for improved efficacy. In addition to reducing computational and memory utilization, this also improves detail capture and long-term reliance. To capture inter-patch interactions while maintaining contextual integrity, the approach applies attention to each partition of the input image. A hybrid self-attention module is constructed by amalgamating conventional and extended rectangular windows within a singular framework, facilitating varied window dimensions for enhanced flexibility and detail retention. This capability augments the model's proficiency in processing various scales and orientations, minimizing generalization mistakes, and enhancing performance for lung cancer detection. Figure 5 illustrates the configuration of these hybrid blocks.

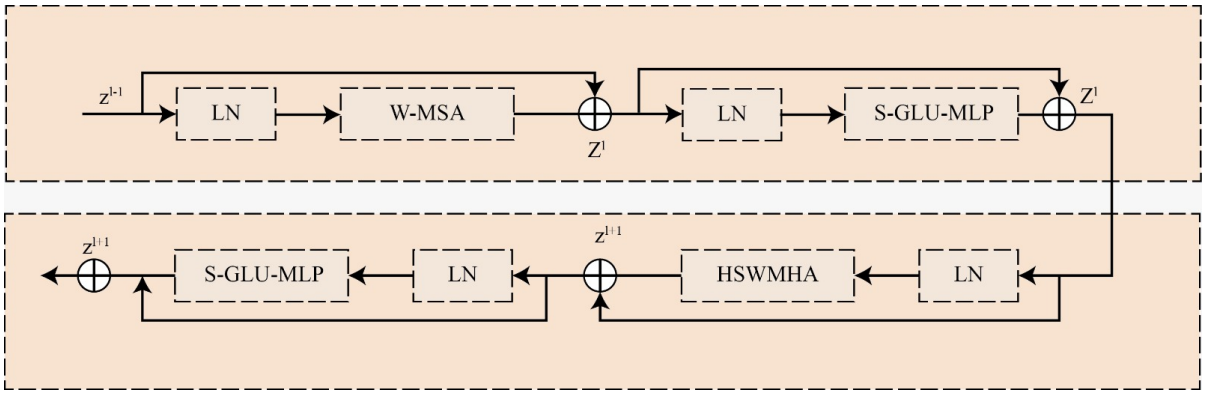

Figure 5. Improved Swin transformer integrating HSWMHA

Figure 5 shows two independent self-attention modules that make up the enhanced transformer blocks. one uses HSWMHA layers, another uses more traditional windows for multi-head self-attention. Initially, it enables the processing of input visuals by conventional shifting window-based self-attention utilizing predetermined window dimensions. In this instance, it performs self-attention on each window to identify local trends. This module uses self-attention stripe windows (horizontal and vertical) to accurately analyze pattern detail across different parts of an image. The use of both horizontal and vertical stripes makes it easier to create long-range relationships, which in turn allows you to gain more comprehensive contextual information. The integration of the three sliding window approaches facilitates the multiple heads of HSWMHA, enabling the model to more effectively comprehend visual input by addressing patterns across various sizes. This strategy is particularly useful for enhancing performance in visual processing tasks.

$$\hat{z}^{l} = W - \mathrm{MHA}(\mathrm{LN}(z^{l-1})) + z^{l-1} \tag{7}$$

$$z^{l} = \mathrm{SGLU\_MLP}(\mathrm{LN}(\hat{z}^{l})) + \hat{z}^{l} \tag{8}$$

$$\hat{z}^{l+1} = \mathrm{HSW\text{-}MSA}(\mathrm{LN}(z^{l})) + z^{l} \tag{9}$$

$$z^{l} = \mathrm{SGLU\_MLP}(\mathrm{LN}(\hat{z}^{l+1})) + \hat{z}^{l+1} \tag{10}$$

Encoder of Swin Transformer

The Swin Transformer's encoder block comprises four sequential stages, as depicted in Figure 2 . Under the window MHA framework, the series of patches is divided into separate windows, where MHA functions independently within each window. Nevertheless, this kind of partitioning results in the loss of border information between windows, namely the absence of connectedness among neighboring windows. In order to address this problem, the attention window is modified to enable the SAM within each window. This leads to a constant reorganization of the original patch sequence during HSWMHA. To restore the original patch sequence order, a reverse loop shift is applied following the HSWMHA process. In specific situations, two consecutive window MHA modules are used instead of alternating between window MHA and HSWMHA modules. This guarantees that the window size in each transformer block is the same as the size of the patch sequence.

Four stages incorporating twelve improved swin transformer blocks make up the transformer encoder module. Layer by layer the window dimension progressively increases across these blocks. Size of the attention window has a big impact on the classification result. We begin with a lesser window size and with C-Swin we gradually combine nearby windows into bigger attention windows. More global characteristics from continuous input tokens can be integrated into the transformer encoder, which promotes improved feature fusion. Moreover, the depth of the model changes the number of attention heads in each window. Every window incorporates different information from several local characteristics; hence this modification is required. Thus, to record a wide range of semantic information and adjust to this changing complexity, it becomes imperative to increase the number of attention heads.

Experimental Results

The following section offers a thorough analysis of the evaluation criteria used to evaluate the suggested technique. It also covers the hardware and software needs for evaluation and model training. Moreover, this section presents comparison studies of the results acquired by using the proposed model.

Dataset and Experimental Setting

The IQ-OTH/NCCD lung cancer datasets, gathered at the Iraq-Oncology Teaching Hospital and the National Center for Cancer Diseases, were used in the experimental study. The data was collected over three months in 2019 and includes lung scans from healthy persons as well as patients with various stages of lung cancer. Oncologists and radiologists created a team to provide expert commentary. Overall, 2073 chest CT scan pictures compose the data, which includes 110 cases in total, reflecting a diverse demographic range with an emphasis on living conditions, educational levels, gender, age, and resident areas. A selection of images from this dataset is illustrated in Figure 1. 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. 40 of the 110 cases were classified as malignant, 55 as normal, and 15 as benign. The Kaggle repository provides public access to the "IQ-OTH/NCCD" dataset.

This research employed the Python programming language and the Keras deep learning framework to conduct the trials. The models were trained on a system equipped with an NVIDIA QUADRO M2000 GPU, 32 GB of RAM, and the Windows 11 operating system. The model underwent training for 100 epochs via the Adam optimiser and cross-entropy loss function, commencing with an initial learning rate of 0.001. The results obtained are presented below, accompanied by relevant analysis and discussion.

Performance Metrics

Evaluation of classification results was based on basic performance metrics (i.e., accuracy; F1-score; precision; and recall), where accuracy is the earliest performance metric calculated in the evaluation of a classifier as it is calculated from the number of accurate predicted samples (i.e., true predictions). Recall is useful as a performance metric in determining how well a classifier will classify unbalanced datasets, while precision is used to provide an indication of the accuracy of a classifier in terms of how many of the items classified have been correctly classified so that it gives a clear indication of the accuracy of the diagnostic rate reported for a given number of true positives. The F1-score was used to combine the effects of both precision and recall based on an additional parameter.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{11}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{12}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{13}$$

$$\text{F1-Score} = \frac{2TP}{2TP+FP+FN} \tag{14}$$

Classification refers to the act of correctly classifying items into one or more classes. Classification is considered a success when an item is classified into its correct category. When this occurs with a sample (which is an individual unit of analysis),

that is referred to as a true positive (TP). When a sample is classified in error as not being a member of its correct category, this is referred to as a false negative (FN). Similarly, when the sample is predicted to be classified as belonging to its category when in reality it does not, this is called a false positive (FP). Finally, a true negative (TN) occurs when a sample is perceived to be classified correctly as not belonging to an appropriate category.

Classification Results

The research utilised the IQ-OTH/NCCD dataset, comprising 3,598 lung CT scan pictures subsequent to the implementation of data augmentation techniques. Among them, 968 were categorised as benign, 1,400 as malignant, and 1,230 as normal. Prior to inputting into the C-Swin model, the images underwent a sequence of preparation procedures. A designated training set and test set were employed to train and assess the model, accordingly. The findings indicated that the test exhibited an accuracy of 96.26% and a precision of 97.48%. Table 1 offers a comprehensive overview of the model's performance, while Figure 6 illustrates the confusion matrix. Figures 7 and 8 illustrate the accuracy and loss graphs for validation and training, respectively.

Table 1. Result of proposed C-Swin model

Class	Precision (%)	Recall (%)	F1-Score (%)
Benign	97.65	95.76	96.32
Malignant	96.38	94.63	97.04
Normal	98.40	98.78	98.90
Accuracy			96.26
Macro Average	97.48	96.39	97.42
Weighted Average	97.35	96.50	97.78

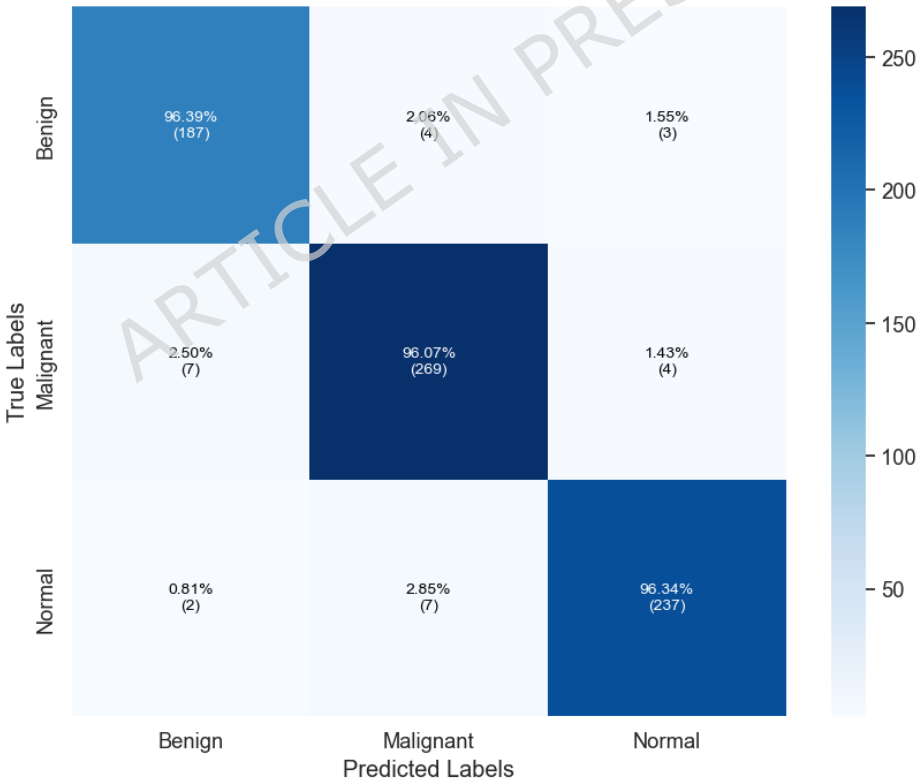

Figure 6. Confusion matrix of the proposed model

A 5-fold cross-validation was performed to test the proposed model's generalizability. Table 2 shows the average output based on the classification results for each fold. Modeled with a mean value accuracy of 96.21%, the precision, recall, and F1-score were 97.33%, 96.39%, and 97.24%, respectively, with a minimal standard deviation across all folds. These findings demonstrate the stability and robustness of the proposed C-Swin model in a variety of data divisions.

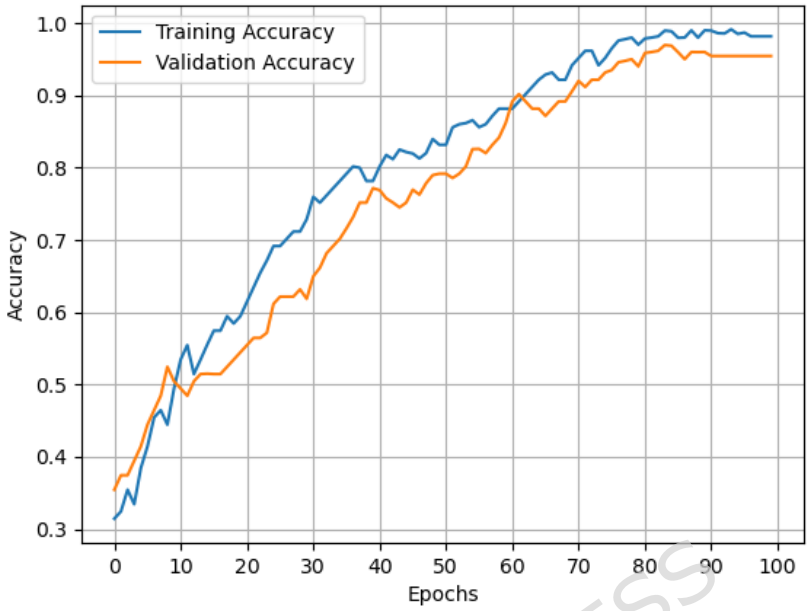

Figure 7. Training and validation accuracy for proposed model

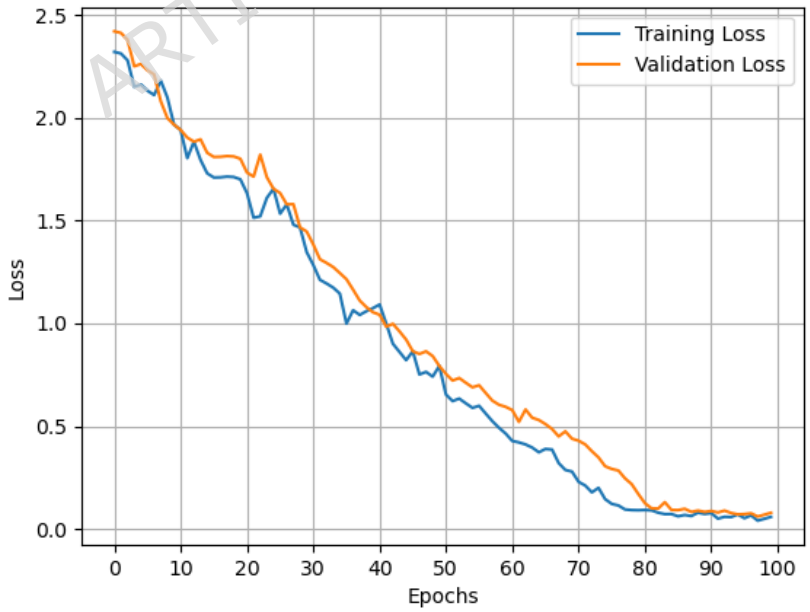

Figure 8. Training and validation loss for proposed model

Table 2. Performance of the proposed C-Swin model using 5-fold cross-validation

Fold	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Fold 1	96.12	97.20	96.25	97.10
Fold 2	96.35	97.50	96.40	97.95
Fold 3	96.04	97.85	96.31	96.94
Fold 4	96.42	97.65	96.52	97.96
Fold 5	96.37	97.20	96.45	97.15
Average	**96.26**	**97.48**	**96.39**	**97.42**
Std. Dev	±0.15	±0.21	±0.10	±0.25

Analysis of Precision-Recall and ROC Curves

The trade-off between precision and recall across three categories–benign, malignant, and normal–is illustrated by the Precision-Recall (PR) curve in Figure 9. The C-Swin model achieved an AP of 0.93 for benign, 0.94 for malignant, and 0.95 for the normal class. This data demonstrates the model's consistently high performance across all classes, with the normal class exhibiting the highest performance and a micro-average precision-recall score of 0.94, indicating the model's efficacy in multi-class classification.

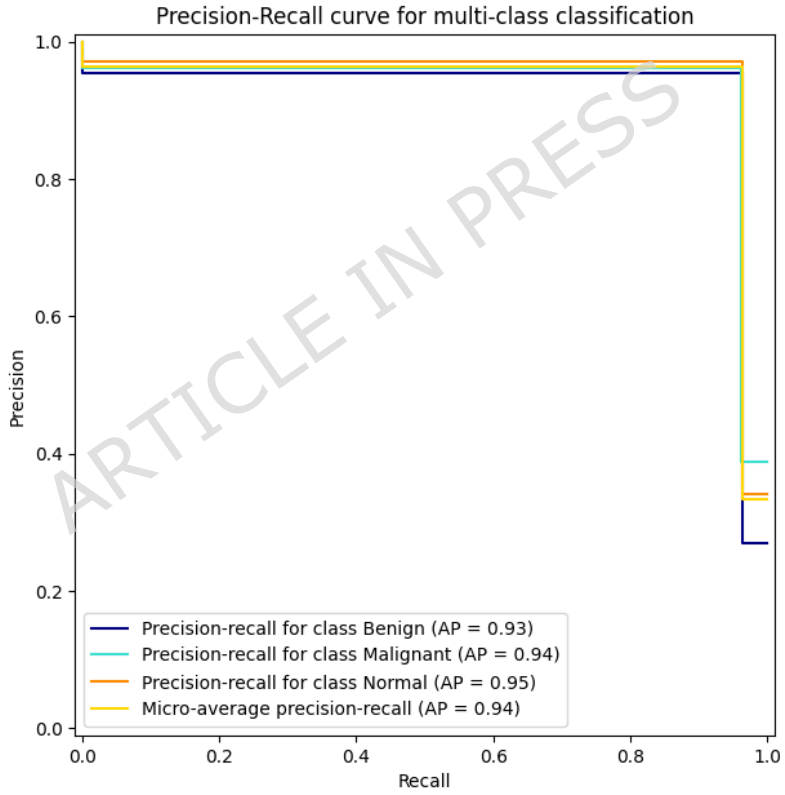

Figure 9. Precision-Recall curve for multi-class classification. The plot shows the Precision-Recall curve for each class (Benign, Malignant, and Normal) along with the Micro-average precision-recall score.

Figure 10 illustrates the Receiver Operating Characteristic (ROC) curve for the C-Swin model. The ROC curve demonstrates the efficacy of the C-Swin model in classifying the categories. The charts illustrate the correlation between True Positive Rate (TPR) and False Positive Rate (FPR) for each of the three distinct classes. The AUC values for each class are as follows: benign 0.93, malignant 0.96, and normal 0.97. Based upon these results, it is clear that the C-Swin model has the most discriminatory power in terms of the normal class and then malignant followed by benign. The diagonal dashed line on the ROC curve demonstrates the baseline randomness that exists for a random classifier, and the C-Swin model is significantly above that line demonstrating its superior ability to classify the classes.

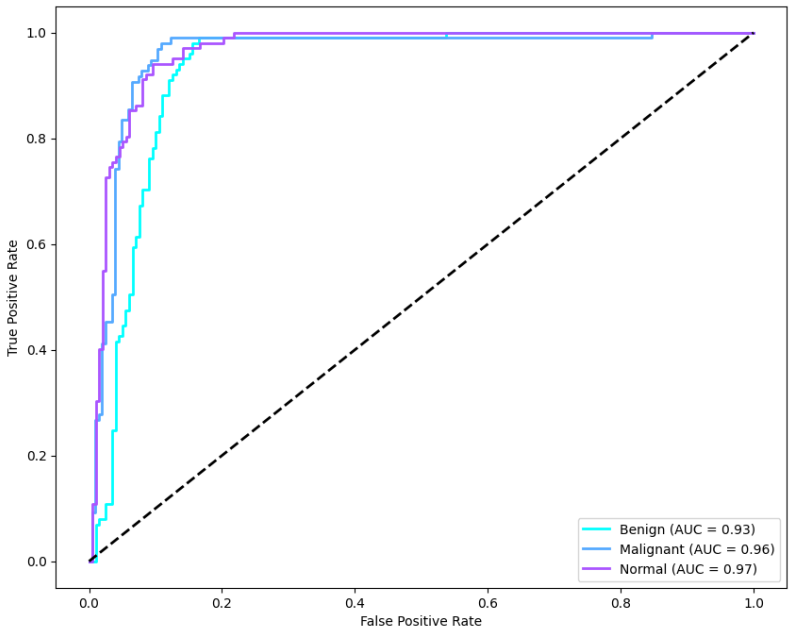

Figure 10. ROC curve for multi-class classification. The plot shows the ROC curve for each class (Benign, Malignant, and Normal) with their corresponding AUC values.

Statistical Test

In order to determine if the performance gains demonstrated by the C-Swin model were statistically significant and assess the stability of the obtained results, formal statistical analyses were performed using the following measurement metrics: accuracy, precision, recall and F1-score. The paired t-test and Wilcoxon signed-rank test were applied to compare the C-Swin with the baseline methods. The paired t-test will show if there is a statistically significant difference in the mean performance values of the proposed method versus the baselines. The Wilcoxon signed-rank test will be used as it is a non-parametric alternative and therefore can be used when distributional assumptions of normality may not be met. Specifically, paired t-tests were used to compare the accuracy and F1-score of each model. The detailed statistical outcomes are presented in Table 3.

Table 3. Paired t-Test Results for Accuracy and F1-Score

Metric	t-Statistic	p-Value
Accuracy	3.45	0.015
F1-Score	4.12	0.008

The result of using a paired t-test for calculation of both the t-value (3.45) and the related p-value (0.015) provides sufficient evidence to reject the null hypothesis, and show that there is a statistically significant improvement in accuracy for the C-Swin model as compared to the baseline model (based on the p-value falling below the significance threshold of 0.05). Also, the t-test for F1-score created a t-statistic (4.12) and p-value (0.008), that proved a statistically significant improvement in F1-score performance as well. In order to confirm these results under situations where the assumption of normality might not be valid, the Wilcoxon signed-rank test was conducted as a non-parametric alternative to the t-test. Both accuracy and F1-score were included in this analysis. The Wilcoxon test results are shown in Table 4.

Table 4. Wilcoxon Signed-Rank Test Results for Accuracy and F1-Score

Metric	W-statistic	p-Value
Accuracy	45	0.023
F1-Score	52	0.014

Results of the Wilcoxon signed-rank test gave a W-value of 45 and a p-value of 0.023 for performance, showing statistical importance. For the F1-score, the test gave a W-score of 52 and a p-value of 0.014, also confirming that the increase in

F1-score is statistically important. The results of both the paired t-test and the Wilcoxon signed-rank test show that there is strong statistical evidence that the C-Swin model is better than baseline methods. The results from both the parametric and non-parametric tests show that the increases in accuracy and the F1-score achieved using the proposed model are strong and consistent. Since all of the p-values obtained from the tests were less than the 0.05 level, it indicates that there is a very small likelihood that the improvements were due to random chance, confirming the overall effectiveness of the proposed C-Swin framework.

Grad-CAM Visualization for Model Explanation

Gradient-weighted Class Activation Mapping (Grad-CAM) elucidates the regions of a CT scan image that significantly influence the decision-making of a ML classification system, so enabling doctors to visualise the model's areas of focus within the image. The Grad-CAM visuals derived from CT scans are categorised into three groups: benign, malignant, and normal, as illustrated in Figure 11. The initial row displays example images depicting CT scans of three conditions. Situated beneath those images, in the second row, are Grad-CAM images that depict the segments of the original scans on which the model focused to determine its predicted categorisation for each scan. Grad-CAM visualisations furnish clinicians with insights into the rationale behind the model's decision. The intensity of the heatmap indicates the level of importance assigned to each region of the CT scan. Brighter areas represent higher relevance.

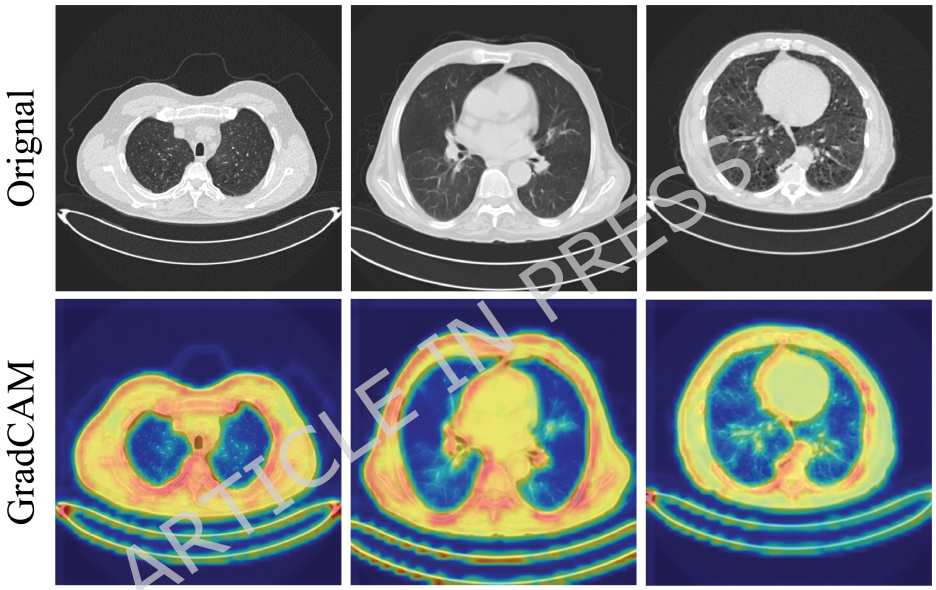

Figure 11. Grad-CAM visualization of lung cancer classification

Ablation Analysis

The ideal mix is selected from various data split ratios to attain the most advantageous results for training purposes. The ideal data split ratio of 70:20:10 surpassed other ratios, achieving an accuracy of 96.26%, precision of 97.48%, recall of 96.39%, and F1-score of 97.42%. The condition exhibiting the lowest accuracy, at 87.25%, along with a precision of 84.98%, recall of 85.11%, and F1-score of 84.49%, corresponds to the ratio of 50:25:25. A comparison of performance across different data partitioning ratios is provided in Table 5 and visualized in Figure 12.

Table 5. Results of varying data split ratios

Split Ratio	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
50:25:25	87.25	84.98	85.11	84.49
60:30:10	91.88	90.10	91.65	91.90
60:20:20	95.54	96.21	94.86	95.67
70:20:10	**96.26**	**97.48**	**96.39**	**97.42**
80:10:10	94.32	93.46	95.31	95.85

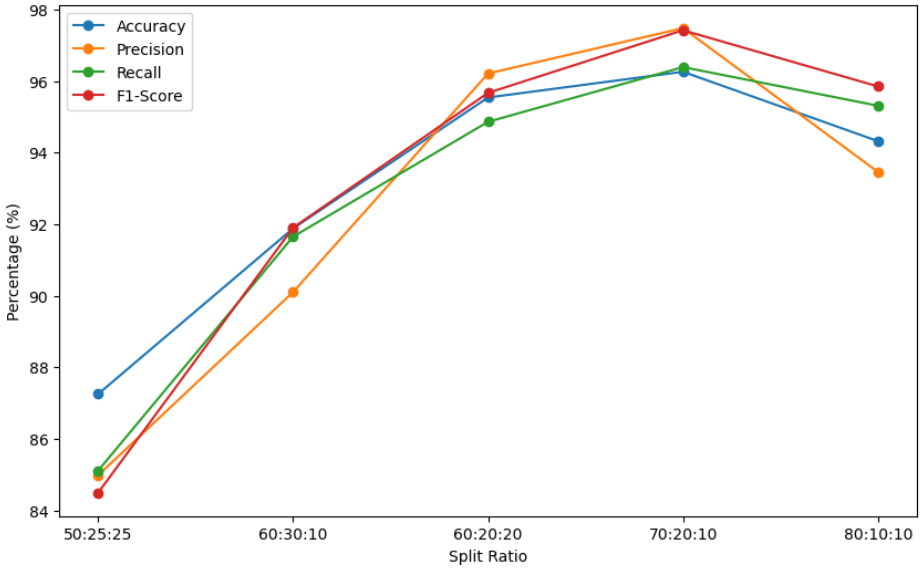

Figure 12. Comparison of accuracy on different data split ratio

A series of experimental comparisons are conducted using pretrained CNN models to assess proposed approach and demonstrate the efficacy of the C-Swin in lung detection and classification. The five pretrained models that we subject to testing are Xception, VGG16, VGG19, InceptionResNetV2, and DenseNet201. Table 6 and Figure 13 compares the outcomes of all models. This demonstrates the absence of a prevalent model in CNN-based methodologies. Each model is more robust in a single respect. Nevertheless, C-Swin is the most advantageous of all criteria. Initially, the C-Swin model exhibits satisfactory performance, achieving the highest average accuracy of 96.26%.

Table 6. Comparison of C-Swin with pre-trained models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Xception[63]	86.45	84.32	83.76	84.00
VGG16[60]	87.89	83.55	85.12	86.33
VGG19[60]	83.22	84.01	83.47	83.73
InceptionResNetV2[64]	84.78	85.24	84.85	85.04
DenseNet201[65]	81.65	80.13	82.68	84.90
C-Swin (Proposed)	**96.26**	**97.48**	**96.39**	**97.42**

To assess the individual contribution of the proposed modules, component-wise ablation of experiments was performed. The greatest results were from the C-Swin model, which included both S-GLU and HSWMHA, with 96.26% accuracy, 97.48% precision, 96.39% recall, and an F1-score of 97.42%. Using regular ReLU instead of S-GLU reduced nonlinearity and dynamic gating, lowering the F1-score to 95.68%. In the same vein, replacing HSWMHA with normal windowed-based MHA resulted in additional degradation, with accuracy dropping to 93.84% and F1-score to 94.50%. These findings, presented in Table 7, demonstrate that each module plays an important role in increasing model classification performance.

Table 7. Component-wise ablation results showing the performance degradation when replacing S-GLU with ReLU and HSWMHA with standard MHA

Model Variant	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
No CNN (Transformer only)	90.12	91.50	90.30	90.70
No Transformer (CNN only)	87.89	83.55	85.12	86.33
w/o S-GLU (ReLU used)	94.81	95.12	94.53	95.68
w/o HSWMHA (Standard MHA)	93.84	94.01	93.47	94.50
C-Swin (Full Model)	**96.26**	**97.48**	**96.39**	**97.42**

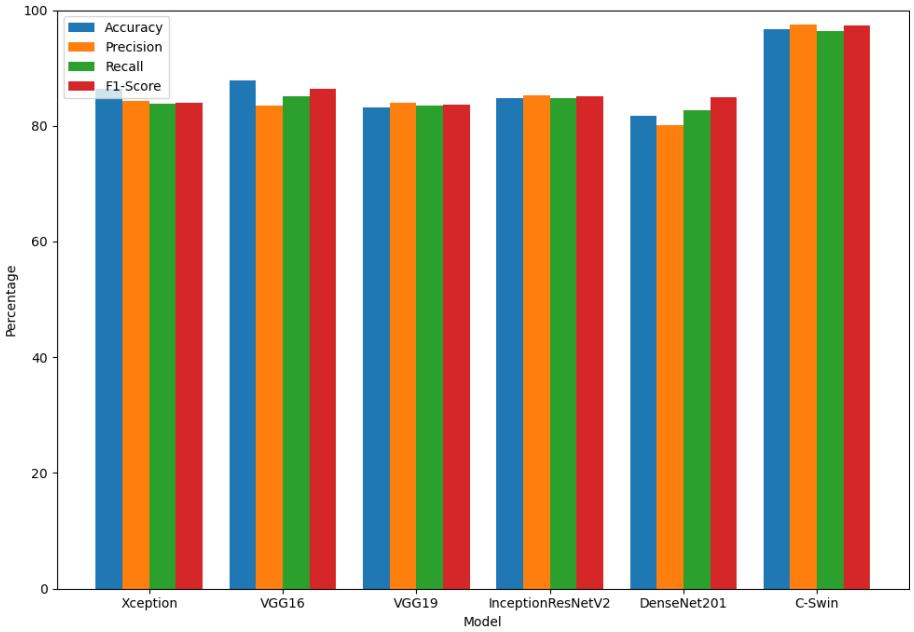

Figure 13. Comparison of evaluation metrics on all models

Table 8 illustrates the extent to which various preprocessing and augmentation approaches affect the performance of the model and demonstrate how they were essential to obtaining optimal results. Without any normalization, the model's accuracy was 94.00%, while precision, recall, and F1-score were 94.58%, 93.88%, and 94.05%, respectively. This indicates that normalization is a key factor in increasing the model's efficiency, particularly regarding precision and recall. However, the accuracy without augmentation was slightly better at 94.12%, with only small differences in precision and recall. Thus, while augmentation has some value, the overall effect of augmentation on the performance of the model is not as great as the effect of normalizing the input data. The greatest decrease in performance is observed when neither normalization nor augmentation is applied, resulting in an accuracy of 92.48% with significant decreases in precision of 93.10%, recall of 92.33%, and F1-score of 92.73%. Therefore, combining both normalization and augmentation techniques is crucial to achieving overall optimal model performance. The highest performance occurs when normalization and augmentation are applied to the model in their entirety (C-Swin Full Model) with an accuracy of 96.26%, along with other metrics at their highest levels, indicating a dramatic improvement in the predictive ability of the model due to the integration of both preprocessing techniques.

Table 8. Model Performance with Different Preprocessing/Augmentation Steps

Preprocessing/Augmentation Step	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
w/o Normalization	94.00	94.58	93.88	94.05
w/o Augmentation	94.12	94.70	93.52	94.07
w/o Normalization & Augmentation	92.48	93.10	92.33	92.73
C-Swin (Full Model)	**96.26**	**97.48**	**96.39**	**97.42**

Table 9 shows the model's results with different optimization algorithms, specifically measuring their effects on accuracy, precision, recall, and F1-score. Stochastic Gradient Descent (SGD) was able to achieve the accuracy of 94.35% amongst the evaluated algorithms, with 94.01% precision, 93.57% recall and 93.88% F1-score. RMSprop slightly improved upon the results achieved by SGD with an accuracy of 94.58% and exhibited improved generalisation over SGD through improved precision of 94.30%, recall of 93.80%, and F1-score of 94.12%. Adagrad performed worse than both SGD and RMSprop, achieving an accuracy of 93.70%, as well as decreased precision, recall and F1-score. Adadelta gauged slightly better than Adagrad 93.90% accurate, however, the Adam optimizer overwhelmingly performed the best amongst the algorithms being tested, achieving the highest accuracy at 96.26%, with the highest precision of 97.48%, recall of 96.39%, and F1-score of 97.42%.

Table 9. Model Performance with Different Optimizers

Optimization Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
SGD	94.35	94.01	93.57	93.88
RMSprop	94.58	94.30	93.80	94.12
Adagrad	93.70	93.20	92.45	92.81
Adadelta	93.90	94.10	93.05	93.57
Adam	**96.26**	**97.48**	**96.39**	**97.42**

Discussion

The proposed C-Swin DL model has an overall accuracy of 96.26%, which is far higher than prior methods for lung cancer detection from CT images. Precision, recall, and F1-score were also excellent for the C-Swin model, with 97.48%, 96.39%, and 97.42%, respectively. This shows that the C-Swin model detects lung cancer status accurately and uses multiple measures to compare its performance to older methods. The hybrid CNN and transformer framework to capture both local and global properties from CT scans, which contributed to its overall effectiveness.

The high performance of the C-Swin model highlights the power of DL in capturing complex patterns in medical imaging data. In addition, utilizing the CNN module for local feature extraction and transformers as global context provides the C-Swin model with an extensive toolset for overcoming the difficulties associated with lung cancer diagnosis. Our findings support the hypothesis that DL models, especially those combining CNNs with transformer-based mechanisms, can substantially improve diagnostic accuracy compared to traditional methods. Moreover, our study is consistent with other research showing convolutional networks can be utilized to discern complex image characteristics and therefore validate the use of those networks in analyzing medical images.

The experimental outcomes shown in Table 10 compared against established state-of-the-art classifiers demonstrate the superior performance of the proposed C-Swin model. The C-Swin model combines CNN and an advanced version of Swin Transformer through its use of hybrid shifted windows to outperform existing approaches with improved accuracies of around 2.31%-6.81% higher than existing methods. This shows that the hybrid approach taken provides a more effective means of capturing spatial relationships between different parts of the image, resulting in higher performance with respect to classifying lung cancer images.

Table 10. Comparison of C-Swin with state-of-the-art techniques

Model	Accuracy (%)
SVM[66]	89.88
ViT[67]	91.93
Ensemble Learning[68]	92.80
AlexNet[58]	93.54
Transfer Learning[57]	94.38
Multi-modal Classification[69]	93.00
Optimized CNN Framework[70]	94.00
C-Swin (Proposed)	**96.26**

The C-Swin model's successful application provides several opportunities in the development of AI-enhanced diagnostic solutions consistent with the principles of medico-technical imaging, particularly regarding the identification of lung cancer at earlier stages due to increased levels of accuracy presented by the model. This can lead to quicker diagnoses thereby facilitating a quicker treatment process resulting in a better outcome for patients. Using both an automated and enhanced imaging solution within everyday clinical processes is now becoming a standard practice in how ML algorithms are applied in the medical imaging industry. Consequently, many radiologists have been able to reduce their workloads and focus on managing increasingly complex case workloads while still adhering to best-practice workflow procedures. The data from this research may encourage more similar AI-assisted diagnostic tools for supporting diagnosis on all types of cancers and other diseases that have not yet been officially diagnosed. The C-Swin model could also be adjusted to allow other types of imaging methods (i.e., MRI or X-ray) to be utilized in developing a more universal diagnostic tool for medical imaging

The ethics surrounding AI use in medicine can greatly affect AI-assisted diagnostic technology's acceptance by clinicians in the clinical setting. Clinician must obtain a patient's consent prior to using a patient's medical record in developing any AI-assisted diagnostic tool using that patient's medical records. In addition, clinician confidence in AI-assisted clinical

decision-making tools can be increased through transparent communication regarding AI-supported clinical decision-making tools and their "explainable" nature through the implementation of XAI techniques like Grad-CAM are examples of ways clinicians are able to understand and confidently utilize AI-supported clinical decision-making tools. In addition, the model must validate the predictive capabilities of the AI-supported clinical decision-making tool using a variety of demographically diverse patients to ensure that the model does not inadvertently create a bias in the AI-supported clinical decision-making process from a demographic category like age, ethnicity, gender, etc. In addition, through external validation using multiple institutions with multiple dataset, the model will be enhanced by confirming that it is generalizable. In addition, through clinical trial design that is multicenter will examine the AI-supported clinical decision-making tool's usability across multiple types of healthcare delivery systems and the use of prospective studies that track usage data of the tool's impact on clinician workflows in real time.

Despite the C-Swin model demonstrating considerable accuracy, it possesses certain limitations. A primary limitation pertains to the amount of the dataset employed for training and validation. Despite the model attaining elevated accuracy on the supplied data, its limited size may hinder its ability to generalise to bigger, more diverse populations. The dataset's acquisition from a single hospital may result in geographic bias. Future study should prioritise the expanding the dataset to encompass a broader range of diverse demographics, hence ensuring the model's efficacy across various groups. A potential avenue for future study is to modify the model for application with alternative medical imaging modalities, such as MRI or X-rays, thereby enhancing its utility in clinical environments. Furthermore, future research would focus on integrating the model into clinical practice by collaborating with hospitals and radiology departments to conduct real-time trials and prospectively validate the outcomes.

Conclusion

The C-Swin model, a hybrid deep learning framework that integrates CNN with an enhanced Swin Transformer that employs Hybrid Shifted Window Multi-Head Attention (HSWMHA) and S-GLU activation, is introduced in this paper. The proposed model combined local and global feature extraction capabilities, it performed better in classification than conventional baselines based on CNN and transformers. Five-fold cross-validation of the IQ-OTH/NCCD dataset supports the obtained results, with the model demonstrating an accuracy of 96.21% and 97.24% in terms of F1-score. The efficiency of the HSWMHA and S-GLU components in improving the model's capacity to capture contextual information and fine-grained lesion details was further validated by ablation analysis. Additionally, ablation research revealed that both the S-GLU and HSWMHA components are effective in enhancing the model's ability to extract the finer features of lesions and contextual information.

Even though the performance is promising, the study has limitations. First, the dataset is comparatively small and originates in one community repository, therefore restraining the generalization of the model to other groups or imaging conditions. Second, although an augmentation of training data was performed, no external validation on a separate dataset was done, which would be critical to establish applicability in the real-world. In the future, we aim to test the model on larger and multi-center datasets, and look into the use of optimized or custom CNN architectures for increased efficiency and deployment in real-time clinical environments. Additionally, incorporating multi-modal data, such as radiomics or genomic data, could enhance the model's diagnostic capabilities and lead to a more comprehensive understanding of lung cancer.

Funding

This research was supported by the research fund of Hanyang University (HY-202500000003701). Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2025R853), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Data availability

The primary data featured in the study can be accessed:
https://www.kaggle.com/datasets/hamdallak/the-iqothnccd-lung-cancer-dataset.
The source code may be found in the Github repository at:
https://github.com/Samia-Nawaz/c-swin-for-lung-cancer-calssification

Competing Interest Declaration

The authors declare no conflicts of interest.

Acknowledgments

This research was supported by the research fund of Hanyang University (HY-202500000003701). The authors extend their appreciation to the Deanship of Research and Graduate Studies at King Khalid University for funding this work through Large Research Project under grant number RGP2/xxx/xx. Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2025R853), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. Ongoing Research Funding program, (ORF-2025-xxx), King Saud University, Riyadh, Saudi Arabia. The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through the project number “NBU-FFR-2024- 2913-xxx".

Author contributions

S.N.Y.: conceptualization, methodology, data Curation, software, visualization, writing – original draft preparation. I.M.N.: conceptualization, methodology, data curation, software, visualization, writing – original draft preparation. S.M.: investigation, resources, writing – review and editing. N.N.: validation, investigation, writing – original draft preparation. A.A.: resources, writing – original draft preparation, visualization. M.A. investigation, resources, writing – review and editing. E.K.: conceptualization, methodology, supervision, funding, writing – review and editing. All authors have read and agreed to the published version of the manuscript.

References

1. Sun, W., Zheng, B. & Qian, W. Computer aided lung cancer diagnosis with deep learning algorithms. In *Medical imaging 2016: computer-aided diagnosis*, vol. 9785, 241–248 (SPIE, 2016).
2. Zhou, Z.-H., Jiang, Y., Yang, Y.-B. & Chen, S.-F. Lung cancer cell identification based on artificial neural network ensembles. *Artif. intelligence medicine* **24**, 25–36 (2002).
3. Nie, L. *et al.* Disease inference from health-related questions via sparse deep learning. *IEEE Transactions on knowledge Data Eng.* **27**, 2107–2119 (2015).
4. Nie, L. *et al.* Beyond doctors: Future health prediction from multimedia and multimodal observations. In *Proceedings of the 23rd ACM international conference on Multimedia*, 591–600 (2015).
5. Dhaware, B. U. & Pise, A. C. Lung cancer detection using bayasein classifier and fcm segmentation. In *2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT)*, 170–174 (IEEE, 2016).
6. Silva, G. L. F. d., Carvalho Filho, A. O. d., Silva, A. C., Paiva, A. C. d. & Gattass, M. Taxonomic indexes for differentiating malignancy of lung nodules on ct images. *Res. on Biomed. Eng.* **32**, 263–272 (2016).
7. Rattan, S., Kaur, S., Kansal, N. & Kaur, J. An optimized lung cancer classification system for computed tomography images. In *2017 Fourth International Conference on Image Information Processing (ICIIP)*, 1–6 (IEEE, 2017).
8. Saba, T., Khan, M. A., Rehman, A. & Marie-Sainte, S. L. Region extraction and classification of skin cancer: A heterogeneous framework of deep cnn features fusion and reduction. *J. medical systems* **43**, 289 (2019).
9. Kuruvilla, J. & Gunavathi, K. Lung cancer classification using neural networks for ct images. *Comput. methods programs biomedicine* **113**, 202–209 (2014).
10. Kumar, D., Wong, A. & Clausi, D. A. Lung nodule classification using deep features in ct images. In *2015 12th conference on computer and robot vision*, 133–138 (IEEE, 2015).
11. Detterbeck, F. C., Boffa, D. J., Kim, A. W. & Tanoue, L. T. The eighth edition lung cancer stage classification. *Chest* **151**, 193–203 (2017).
12. Ayshath Thabsheera, A., Thasleema, T. & Rajesh, R. Lung cancer detection using ct scan images: A review on various image processing techniques. *Data Anal. Learn. Proc. DAL 2018* 413–419 (2018).
13. Lakshmanaprabu, S., Mohanty, S. N., Shankar, K., Arunkumar, N. & Ramirez, G. Optimal deep learning model for classification of lung cancer on ct images. *Futur. Gener. Comput. Syst.* **92**, 374–382 (2019).
14. Khan, S. A. *et al.* Lungs nodule detection framework from computed tomography images using support vector machine. *Microsc. research technique* **82**, 1256–1266 (2019).
15. Vasanthi, K. & Kumar, N. B. An efficient lung image classification using gda based feature reduction and tree classifier. In *Handbook of Multimedia Information Security: Techniques and Applications*, 645–666 (Springer, 2019).

16. Nasrullah, N. *et al.* Automated lung nodule detection and classification using deep learning combined with multiple strategies. *Sensors* **19**, 3722 (2019).

17. Ali, A., Shahbaz, H. & Damaševičius, R. xcvit: Improved vision transformer network with fusion of cnn and xception for skin disease recognition with explainable ai. *Comput. Mater. & Continua* **83** (2025).

18. Nasir, I. M. *et al.* An optimized approach for breast cancer classification for histopathological images based on hybrid feature set. *Curr. Med. Imaging Rev.* **17**, 136–147 (2021).

19. Nasir, I. M., Alrasheedi, M. A. & Alreshidi, N. A. Mfan: Multi-feature attention network for breast cancer classification. *Mathematics* **12**, 3639 (2024).

20. Yousafzai, S. N., Nasir, I. M., Tehsin, S., Fitriyani, N. L. & Syafrudin, M. Fltrans-net: Transformer-based feature learning network for wheat head detection. *Comput. Electron. Agric.* **229**, 109706 (2025).

21. Yousafzai, S. N. *et al.* Multi-stage neural network-based ensemble learning approach for wheat leaf disease classification. *IEEE Access* (2025).

22. Yousafzai, S. N. *et al.* Advanced clustering and transfer learning based approach for rice leaf disease segmentation and classification. *PeerJ Comput. Sci.* **11**, e3018 (2025).

23. Alzaidi, M. S. A. *et al.* An efficient fusion network for fake news classification. *Mathematics* **12**, 3294 (2024).

24. Alqadi, B. S. *et al.* Transfer learning driven fake news detection and classification using large language models. *Sci. Reports* **15**, 28490 (2025).

25. Ali, A. *et al.* Towards improved fake news detection using a hybrid roberta and metadata enhanced xgboost model. *Sci. Reports* (2025).

26. Toor, M. S. *et al.* An optimized weighted-voting-based ensemble learning approach for fake news classification. *Mathematics* **13**, 449 (2025).

27. Fernandes, S. L., Rajinikanth, V. & Kadry, S. A hybrid framework to evaluate breast abnormality using infrared thermal images. *IEEE Consumer Electron. Mag.* **8**, 31–36 (2019).

28. Acharya, U. R. *et al.* Automated detection of alzheimer's disease using brain mri images–a study with various feature extraction techniques. *J. medical systems* **43**, 302 (2019).

29. Amin, J., Sharif, M., Yasmin, M. & Fernandes, S. L. Big data analysis for brain tumor detection: Deep convolutional neural networks. *Futur. Gener. Comput. Syst.* **87**, 290–297 (2018).

30. Liaqat, A. *et al.* Automated ulcer and bleeding classification from wce images using multiple features fusion and selection. *J. Mech. Medicine Biol.* **18**, 1850038 (2018).

31. Rajinikanth, V., Satapathy, S. C., Fernandes, S. L. & Nachiappan, S. Entropy based segmentation of tumor from brain mr images–a study with teaching learning based optimization. *Pattern Recognit. Lett.* **94**, 87–95 (2017).

32. Ranjan, R., Arya, R., Fernandes, S. L., Sravya, E. & Jain, V. A fuzzy neural network approach for automatic k-complex detection in sleep eeg signal. *Pattern Recognit. Lett.* **115**, 74–83 (2018).

33. Satapathy, S. C., Fernandes, S. L. & Lin, H. Stroke lesion segmentation and analysis using entropy/otsu's function–a study with social group optimization. *Curr. Bioinforma.* **14**, 305–313 (2019).

34. Wound, I. S. Shannon's entropy and watershed algorithm based technique to inspect. In *Smart Intelligent Computing and Applications: Proceedings of the Second International Conference on SCI 2018, Volume 2*, vol. 105, 23 (Springer, 2018).

35. Raja, N. S. M., Fernandes, S. L., Dey, N., Satapathy, S. C. & Rajinikanth, V. Contrast enhanced medical mri evaluation using tsallis entropy and region growing segmentation. *J. Ambient Intell. Humaniz. Comput.* **15**, 961–972 (2024).

36. Naqi, S., Sharif, M., Yasmin, M. & Fernandes, S. L. Lung nodule detection using polygon approximation and hybrid features from ct images. *Curr. Med. Imaging* **14**, 108–117 (2018).

37. Sakshiwala & Singh, M. P. A new framework for multi-scale cnn-based malignancy classification of pulmonary lung nodules. *J. Ambient Intell. Humaniz. Comput.* **14**, 4675–4683 (2023).

38. Xu, X. *et al.* Multi-scale supervised contrastive learning for benign-malignant classification of pulmonary nodules in chest ct scans. In *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*, 1–4 (IEEE, 2023).

39. Lima, T., Luz, D., Oseas, A., Veras, R. & Araújo, F. Automatic classification of pulmonary nodules in computed tomography images using pre-trained networks and bag of features. *Multimed. tools applications* **82**, 42977–42993 (2023).

40. Roy, R., Mazumdar, S. & Chowdhury, A. S. Adgan: Attribute-driven generative adversarial network for synthesis and multiclass classification of pulmonary nodules. *IEEE Transactions on Neural Networks Learn. Syst.* **35**, 2484–2495 (2022).

41. Ghosal, S. S., Sarkar, I. & El Hallaoui, I. Lung nodule classification using convolutional autoencoder and clustering augmented learning method (calm). In *HSDM@ WSDM*, 19–26 (2020).

42. Singh, D. P., Banerjee, T., Kour, P., Swain, D. & Narayan, Y. Cicada (ucx): A novel approach for automated breast cancer classification through aggressiveness delineation. *Comput. Biol. Chem.* **115**, 108368 (2025).

43. Singh, D. P., Kour, P., Banerjee, T. & Swain, D. A comprehensive review of various machine learning and deep learning models for anti-cancer drug response prediction: Comparative analysis with existing state of the art methods. *Arch. Comput. Methods Eng.* 1–25 (2025).

44. Banerjee, T. *et al.* A novel hybrid deep learning approach combining deep feature attention and statistical validation for enhanced thyroid ultrasound segmentation. *Sci. Reports* **15**, 27207 (2025).

45. Banerjee, T. Electromagnetic interaction algorithm (eia)-based feature selection with adaptive kernel attention network (akattnet) for autism spectrum disorder classification. *Int. J. Dev. Neurosci.* **85**, e70034 (2025).

46. Banerjee, T. *et al.* Pyramidal attention-based t network for brain tumor classification: a comprehensive analysis of transfer learning approaches for clinically reliable and reliable ai hybrid approaches. *Sci. Reports* **15**, 28669 (2025).

47. Singh, D. P. *et al.* A comprehensive study on deep learning models for the detection of diabetic retinopathy using pathological images. *Arch. Comput. Methods Eng.* 1–30 (2025).

48. Banerjee, T. *et al.* A novel unified inception-u-net hybrid gravitational optimization model (uigo) incorporating automated medical image segmentation and feature selection for liver tumor detection. *Sci. Reports* **15**, 29908 (2025).

49. Banerjee, T. Comparing bipartite convoluted and attention-driven methods for skin cancer detection: A review of explainable ai and transfer learning strategies. *Arch. Comput. Methods Eng.* 1–25 (2025).

50. Yousafzai, S. N., Nasir, I. M., Tehsin, S. & Khan, J. A. Mra-net: Multiscale residual attention network for multiclass alzheimer disease classification. In *2024 5th International Conference on Innovative Computing (ICIC)*, 1–8 (IEEE, 2024).

51. Thaljaoui, A. *et al.* Explainable skin cancer diagnosis with parallel attention mechanism for segmentation and classification. *Biomed. Signal Process. Control.* **113**, 109159 (2026).

52. Mukherjee, P. *et al.* A shallow convolutional neural network predicts prognosis of lung cancer patients in multi-institutional computed tomography image datasets. *Nat. machine intelligence* **2**, 274–282 (2020).

53. Sun, R., Pang, Y. & Li, W. Efficient lung cancer image classification and segmentation algorithm based on an improved swin transformer. *Electronics* **12**, 1024 (2023).

54. Chen, J., Ma, Q. & Wang, W. A lung cancer detection system based on convolutional neural networks and natural language processing. In *2021 2nd International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT)*, 354–359 (IEEE, 2021).

55. Yousafzai, S. N. *et al.* A fusion framework of transformer and cnn for non-small cell lung cancer classification. In *International Conference on Smart Systems and Emerging Technologies*, 162–173 (Springer, 2024).

56. Bangare, S. L. *et al.* Computer-aided lung cancer detection and classification of ct images using convolutional neural network. In *Computer Vision and Internet of Things*, 247–262 (Chapman and Hall/CRC, 2022).

57. Al-Huseiny, M. *et al.* Transfer learning with googlenet for detection of lung cancer. *Indonesian J. Electr. Eng. computer science* (2021).

58. Al-Yasriy, H. F., Al-Husieny, M. S., Mohsen, F. Y., Khalil, E. A. & Hassan, Z. S. Diagnosis of lung cancer based on ct scans using cnn. In *IOP conference series: materials science and engineering*, vol. 928, 022035 (IOP Publishing, 2020).

59. Banerjee, T. Towards automated and reliable lung cancer detection in histopathological images using dy-fspan: A feature-summarized pyramidal attention network for explainable ai. *Comput. Biol. Chem.* 108500 (2025).

60. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

61. Kaiser, L. *et al.* One model to learn them all. *arXiv preprint arXiv:1706.05137* (2017).

62. Liu, Z. *et al.* Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, 10012–10022 (2021).

63. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1251–1258 (2017).

64. Szegedy, C., Ioffe, S., Vanhoucke, V. & Alemi, A. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proceedings of the AAAI conference on artificial intelligence*, vol. 31 (2017).

65. Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K. Q. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4700–4708 (2017).

66. Kareem, H. F., AL-Husieny, M. S., Mohsen, F. Y., Khalil, E. A. & Hassan, Z. S. Evaluation of svm performance in the detection of lung cancer in marked ct scan dataset. *Indonesian J. Electr. Eng. Comput. Sci.* **21**, 1731 (2021).

67. Malaviya, N., Rahevar, M., Virani, A., Ganatra, A. & Bhuva, K. Lvit: Vision transformer for lung cancer detection. In *2023 International Conference on Artificial Intelligence and Smart Communication (AISC)*, 93–98 (IEEE, 2023).

68. Solyman, S. & Schwenker, F. Lung tumor detection and recognition using deep convolutional neural networks. In *Pan African Conference on Artificial Intelligence*, 79–91 (Springer, 2022).

69. Uddin, A. H. *et al.* Colon and lung cancer classification from multi-modal images using resilient and efficient neural network architectures. *Heliyon* **10** (2024).

70. Inbasakaran, G. & Ruth, J. A. Clinical-ready cnn framework for lung cancer classification: Systematic optimization for healthcare deployment with enhanced computational efficiency. *Intell. Medicine* 100292 (2025).