A Comparative Analysis of Transformer Architectures for Automated Lung Cancer Detection in CT Images

Baris Okmen[1], Yigitcan Cakmak[1,*], Ishak Pacal[1]

[1]   Department of Computer Engineering, Faculty of Engineering, Igdir Universty, 76000, Igdir, Turkey

ARTICLE INFO	ABSTRACT
               	The imperative for early-stage lung cancer detection is widely recognized as a critical determinant of therapeutic efficacy and patient survival. Conventional diagnostic workflows, however, are frequently constrained by their labour-intensive nature and susceptibility to interpretive inaccuracies, positioning artificial intelligence (AI) as a transformative technology in medical imaging. This research conducts a rigorous comparative analysis of four prominent vision transformer (ViT) architectures; Swin-Base, ViT-Base, DeiT-Base, and BEiT-Base evaluating their performance in the automated classification of lung cancer from computed tomography (CT) scans. The empirical validation was performed on the open-access IQ-OTH/NCCD dataset, a corpus of 1,097 images distributed across benign (n=120), malignant (n=561), and normal (n=416) classes. Model proficiency was quantified using established metrics of accuracy, precision, recall, and F1-score. The findings unequivocally establish the superiority of the Swin-Base model, which, by utilizing its innovative hierarchical design and shifted-window mechanism, attained a benchmark accuracy of 98.80% and an F1-score of 97.52%. While its counterparts achieved commendable accuracies ViT-Base (95.18%), DeiT-Base (96.39%), and BEiT-Base (95.78%) they did not match the performance of Swin-Base. Notably, this leading performance was achieved with greater computational efficiency, requiring a lower GFLOPS count than competing models.

1. Introduction

One of the major causes of cancer deaths is lung cancer, with an estimated one-fifth of cancer deaths being attributed to this disease [1–3]. The pathological characteristics of lung cancer consist of an increase in the number of carcinogenic growths (Tumors) occurring within and along the sides of lung tissue (lung parenchyma). Currently, lung cancer is a worldwide health problem, and according to the World Health Organisation (WHO), each year millions of people die because of lung cancer; however, what is more disturbing is that around 80% of the people diagnosed with lung cancer will experience a poor outcome because they are diagnosed when their tumour has developed to the

---

point that surgical resection is no longer an option [4,5]. Therefore, the need for early and accurate detection of lung cancer is critical because many patients with lung cancer do not exhibit symptoms until the disease is advanced to a fatal stage [6,7]. The detection of lung cancer is primarily accomplished through diagnostic imaging techniques [8], and the current method for interpretation of imaging studies is by manual review of the studies by trained medical professionals, which is a tedious task that requires time and cognitive effort and is subject to errors caused by human judgment. Of the currently available imaging technologies, low-dose computed tomography (LDCT) has become the standard imaging modality, providing the most accurate visualization of the anatomy of the chest and the only modality that can identify many small early-stage lung tumors.

Efforts to alleviate the requirement to interpret the results of CAD systems using the same methodology as the original healthcare provider created CAD systems using traditional computer vision and machine learning techniques [9]. The effectiveness of these experimental systems was greatly diminished because of a major roadblock; the systems were limited by the fact that the systems were developed with "manual" or "crafted" features that greatly decreased the diagnostic accuracy of the systems. To overcome this roadblock, the shift from traditional CAD systems to advanced AI, such as Deep Learning, and, most recently, Transformer models, represents a major shift in the CAD landscape [10]. Deep Learning architectures are the first to provide an end-to-end solution, learning the most relevant features from the input data, rather than relying on manually defined features. Innovating upon this, Transformer-based architecture has rapidly emerged as a very strong alternative to foundational convolutional designs, demonstrating superior modelling of long-range dependencies [11]. In sharp contrast with CNNs that process information via local receptive fields, Transformers leverage a self-attention mechanism for a holistic analysis, weighing concurrently relationships between all constituent parts of an input image. The result is a much more contextually aware and meaningful feature extraction process, hence accelerating the adoption of active end-to-end learning frameworks for high-stakes classification and segmentation tasks in medical computer vision.

Imaging methods are already in extensive use in many medical subfields, including lung and blood diseases, for diagnosing diseases [12–14]. Todays, AI-based early diagnosis systems are used to detect numerous health issues using medical images acquired using the most widely used imaging methods [15]. These systems have already been used to detect various conditions, including retinal diseases [16,17], breast cancer [18–25], cervical cancer [26–29], and brain tumors [30–33]. In this study, we evaulte to utilize four different widely used transformer architectures Swin-Base, ViT-Base, DeiT-Base, and DeiT-Base to classify lung cancer lesions, which contained CT (Computed Tomography) images. Subsequently, we performed a thorough classification process by using these transformer-based models and reviewed the results in detail according to performance, accuracy, and overall success criteria.

2. RELATED WORKS

Transformers are showing promise as source of improvement to traditional imaging technologies. Using hybrid neural networks of CNNs and transformers has begun to produce better model performance through improved use of both convolutional feature extraction and transformer global context modelling. To cite a few examples of such developments, Debnath et al. [34] developed a new hybrid model called Lung MobileVIT, which incorporates depthwise separable convolutions along with self-attention and uses attention mechanisms such as CBAM to amplify discriminative power, especially when trained and tested on unbalanced datasets. Likewise, Yogappan et al. [35] developed a new T-RGB algorithm, which uses a preliminary CNN approach for extracting accurately layered feature maps and then uses these within a Vision Transformer framework for classification, employing

a new bio-inspired Gooseneck Barnacle strategy for optimizing hyperparameters. To overcome the high computational requirements associated with traditional Transformer models, a new lightweight Vision Transformer named LwViT-DL was developed by Mannepalli et al. [36], which employs a conditional variational auto-encoder based feature extraction strategy and a new custom dual-attention Vision Transformer model framework.

Transformers have evolved beyond simple classification tasks to include more sophisticated target diagnostic problems, e.g., segmentation, detection (diagnosis), and multi-modal risk prediction. With respect to lung nodule detection, Tang et al. [37] proposed (using) the DEtection TRansformer, LN-DETR. This model was specifically designed to enhance the performance of identifying small and morphologically complex lung nodules by adding specialized modules for multi-scale attention and cross-scale feature fusion. In a similar fashion, Cui et al. [38] introduced SF2T, a novel two-stage framework using a 3D Swin Transformer to generate candidate nodules, and a new two-stream network to reduce false positives, which demonstrates the advantages of advanced variations of Transformers for challenging diagnostic detection tasks. Furthermore, the several applications of these architectures have been demonstrated by Li et al. [39] with the use of a multimodal Swin Transformer that fuses CT imaging and pathological data to predict the risk of bone metastasis. Additionally, Talib et al. [40] address the complete diagnostic process by proposing specialized models for segmentation (TransSegNet) and classification (MinClassNet), highlighting that diagnostic systems are now designed as fully automated end-to-end solutions.

3. MATERIALS AND METHODS

3.1 Dataset

The research used the open and public Iraq-Oncology Teaching Hospital/National Center for Cancer Diseases (IQ-OTH/NCCD) dataset (10.1007/s11270-022-05875-0); this dataset was gathered over three months in late 2019. The dataset included Computed Tomography (CT) imaging of 110 patients, including healthy patients and patients of stages I-IV lung cancer. All images were labeled by appropriate oncologists and radiologists who carefully labeled the image stack. The study was approved by Research Ethics Boards, where all identifying patient information was stripped from all images [41,42].

CT scans were taken by this study through utilization of a Siemens SOMATOM scanner and according to a used a standard clinical protocol for all CT scan acquisitions. As per the clinical protocol, scans were undertaken at 120 kV tube voltage and at a thickness of (1 mm) slice thickness with patients required to be in a full inspiration cycle, as to provide consistent scanning results. The final dataset went through a process of curation, resulting in 1,097 total CT scans that were subsequently divided into three separate diagnostic categories: benign (n=120), malignant (n=561) and normal (n=416). One of the unique features of this dataset is its significant imbalance, with malignant cases being by far the highest in terms of volume with benign cases being considerably lower in volume. Therefore, in order to reduce the chances that models created from this dataset reflect any kind of modelling bias, and to create a more robust evaluation framework, stratified sampling was used to partition the dataset so as to create separate training (n=767, 70%), validation (n=164, 15%) and testing (n=166, 15%) data subsets while maintaining the original proportion of each of the three classes in the newly created data subsets. A more detailed breakdown of the distributions for each of the three classes is located in Table 1.

Table 1.

Distribution of images in the IQ-OTH/NCCD dataset across training, validation, and test sets.

Class	Training	Validation	Test	Total

Benign	84	18	18	120
Malignant	392	84	85	561
Normal	291	62	63	416
Total	767	164	166	1097

In this study, Fig. 1. showcases a variety of axial CT scan slices representing the three diagnostic classes. the images in the top row are classified as benign, the images in the middle row show malignant tumors, and the images in the bottom row are examples of normal lung scans. Each of these image categories represents the differences among classes while at the same time indicating that some features are not as easily recognized or distinguished. Therefore, helping researchers better understand what is needed from a diagnostic model (group) by providing visual examples will ultimately help with the difficulty of this classification challenge.
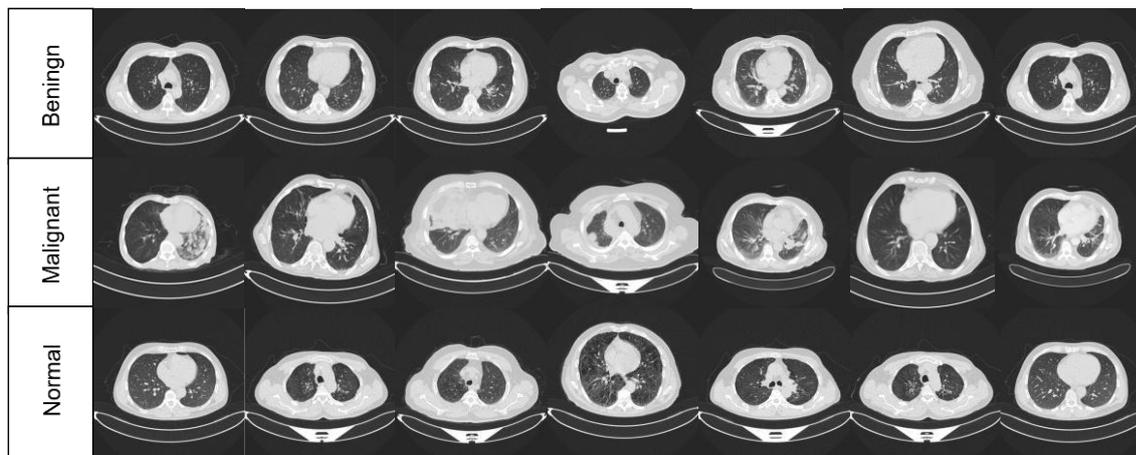

Fig. 1. Representative sample images from the IQ-OTH/NCCD dataset.

3.2 Data Augmentation

To address the common issue of overfitting in AI models and enhance their generalization capabilities, particularly on a limited medical imaging dataset, a range of data augmentation techniques was applied dynamically during the training phase. The augmentation pipeline for each training image included a RandomResizedCrop operation, which randomly selected a portion of the image (scaling between 8% and 100% of the original area with an aspect ratio of 0.75 to 1.33) and resized it to a standard 224x224 pixel dimension. This was supplemented by random horizontal flipping (with a 50% probability) and color jittering, which altered the brightness, contrast, saturation, and hue by a factor of 0.4. Notably, vertical flipping was deliberately omitted from this process. This on-the-fly augmentation strategy ensured that the model was exposed to a diverse range of transformed data in each training epoch, fostering the development of more robust and generalizable features [43,44].

3.3 Model Architecture

A new dawn has emerged with the recent progress made in the area of machine learning algorithms, especially with the introduction of the revolutionary "Transformer" approach originally intended for natural language processing tasks into the realm of computer vision as well. ViT (Vision Transformer) [45] is perhaps the leading example within this area, which processes image data by processing them not directly as a feed into a neural network but by processing them as a sequence consisting of "patches" treated as "tokens." This can fully process the image by developing a "self-

attention" mechanism applied on all patches within the image as a whole, without needing prior convolutional processing as a precursor to feed into a neural network as is the need with convolutional networks themselves. DeiT (Data-efficient Image Transformers) [46] is a learning model intended for upgrading the level of ViT by applying a "teacher/student" approach with the aim of either equaling or even outperforming ViT's original level with a significantly reduced dataset size.

To address the computational complexity and scalability of Vision Transformer (ViT) models, Swin Transformer [47] was developed that takes a hierarchical approach. Unlike ViT which applies multi-head self-attention globally to images, Swin Transformer first divides images into smaller "windows," and applies the attention only locally to these windows. This reduced computational cost of local attention and increased it to gaining more of a hierarchical feature representation similar to convolutional networks. The windows are shifted so that they'll interact with one another to capture a wider context. The Swin Transformer's performance has been shown to outperform ViT on more complex computer vision tasks like object detection and semantic segmentation in addition to image classification. BEiT (Bidirectional Encoder representations from Image Transformers) [48] is a Transformer model developed with a self-supervised pre-training task that is similar to successful BERT work in natural language processing. For example, BEiT can predict masked image patches, and thus learn semantic knowledge and structural relationships within images beforehand and then use these rich feature representation templates for further tasks. Each represents the next generation to approaches to image processing.

4. Results and Discussion

Table 2.

Performance metrics of the analyzed transformer models.

Models	Accuracy	Precision	Recall	F1-score	Params	GFLOPs
Swin-Base	98.80	98.97	96.30	97.52	86.75	30.3375
ViT-Base	95.18	96.23	87.97	90.92	85.8	33.7257
Deit-Base	96.39	92.86	92.86	92.86	85.8	33.7257
Beit-Base	95.78	92.03	91.01	91.50	85.76	25.3294

Within this research, four different Transformer-based models were evaluated including the Swin Transformer (Swin-Base), Vision Transformer (ViT-Base), Data-efficient Image Transformer (DeiT-Base), and Bidirectional Encoder representations from Image Transformers (BEiT-Base) with respect to lung cancer diagnosis (Table 2). With regards to the model performance statistics, the Swin-Base model provided by far the best results. The Swin-Base model achieved the highest performance with an accuracy of 98.80% and an F1-score of 97.52%. The main reason for the superior performance of Swin-Base, despite its original introduction to computer vision for a very different application in autonomous driving, derives from its hierarchical architecture and shifted-window mechanism. Architecture enables the Swin Transformer to more effectively capture both local features and long-range context, which helps exploit the complex and nuanced details that are typically required for

classifying medical images. The confusion matrices and ROC curves for the top two models, which demonstrated superior performance, are presented on Fig. 2.
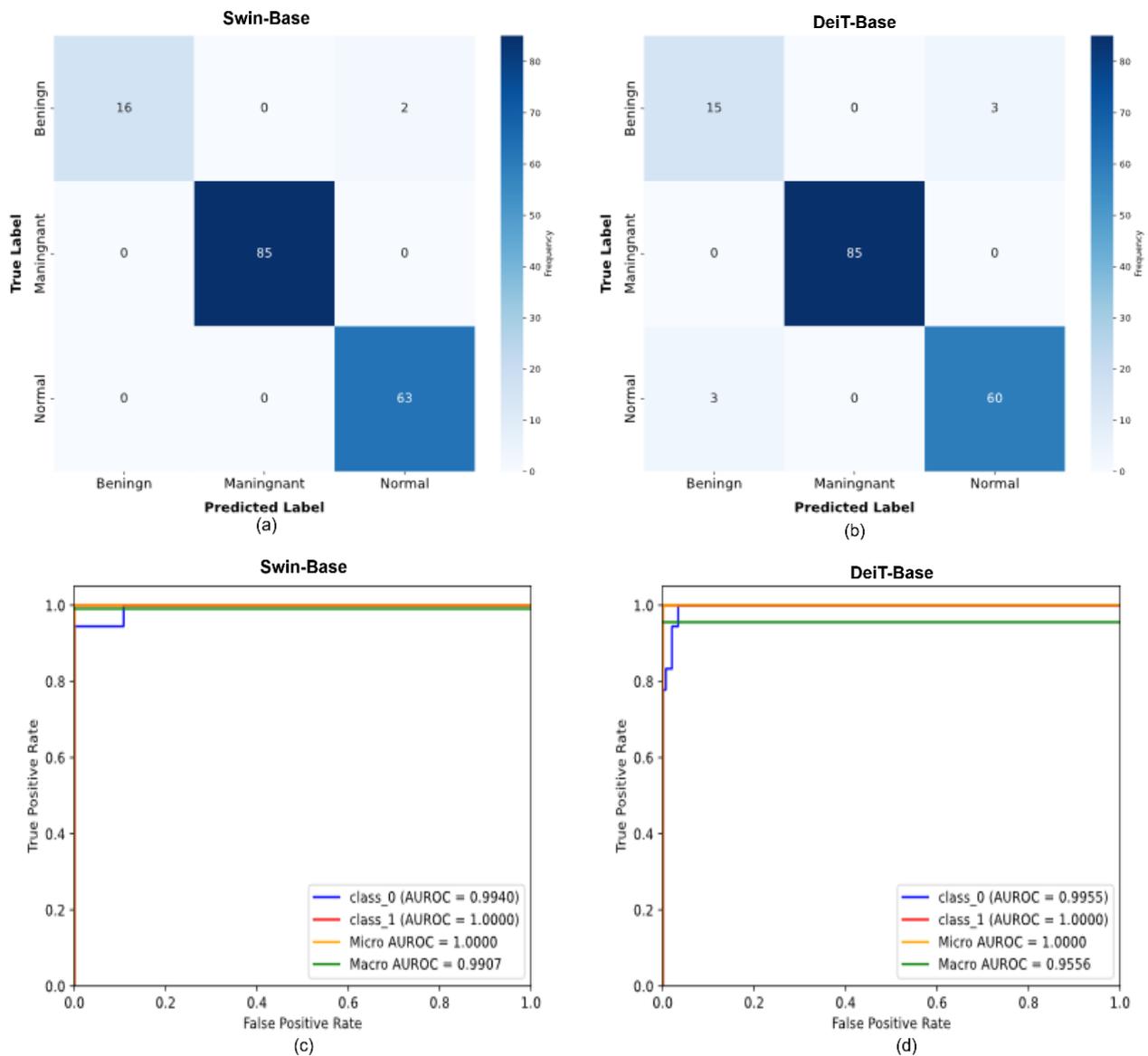

Fig. 2. Performance metrics of the top-performing Swin-Base and DeiT-Base models: (a, b) Confusion Matrices and (c, d) ROC Curves.

When we compare the other models, we can see that DeiT-Base (96.39% accuracy) and BEiT-Base (95.78% accuracy) outperformed ViT-Base (95.18% accuracy) just a little. DeiT outperformed ViT largely because it employs a data-efficient teacher/student learning mechanism across three tasks, while BEiT outperformed ViT broadly because the self-supervised pretext task (masked image modeling) better facilitated the structural understanding of semantics in images. The limited performance of ViT-Base may also indicate additional limitations related to the scale of model reliance on large datasets, particularly when used on a smaller, domain-specific medical image dataset, given that the model parameters (Params) of all models are relatively similar, the GFLOPs value of Swin-Base is lower than that of ViT and DeiT, and the Swin Transformer achieved the best results with a

lower computational burden. As such, Swin is potentially a more cost-efficient and clinically practical model.

5. Conclusions

This study thoroughly evaluated and classified the testing datasets with four different transformer architectures (Swin-Base, ViT-Base, DeiT-Base, and BEiT-Base) with real-world example CT images of lung cancer patients. The results clearly showed that the Swin-Base model (98.80% Accuracy and an F1 score of 97.52%) had better performance than all other models. The models' performance could be due to the Swin's hierarchical model and shifted-window capabilities, which allows the Swin Transformer to model local and long-range complex dependencies in medical images. Moreover, the Swin-Base model provided the best performance with a much lower computational cost (GFLOPS) than ViT and DeiT for clinical settings. Although the ViT-Base model produced the lowest accuracy for all models in classification, the DeiT-Base and BEiT-Base models produced better accuracy because of advantages that support data efficiency and self-supervised learning approaches. In conclusion, we have established that the Swin Transformer can be useful for very complex diagnostic tasks, especially where we have very few medical datasets and we would prefer the Swin Transform in future research studies. Subsequent work could focus on further enhancing its performance by integrating it with various data augmentation techniques and hybrid approaches.

Author Contributions

Conceptualization, B.O., Y.C. and I.P.; methodology, B.O. and Y.C.; software, B.O. and Y.C.; validation, B.O., Y.C. and I.P.; formal analysis, I.P.; investigation, B.O. and Y.C.; resources, I.P.; data curation, B.O. and Y.C.; writing—original draft preparation, B.O. and Y.C.; writing—review and editing, I.P.; visualization, B.O. and Y.C.; supervision, I.P.; project administration, I.P.; funding acquisition, I.P. All authors have read and agreed to the published version of the manuscript.

Funding

Data Availability Statement

The data presented in this study are publicly available. This research utilized the IQ-OTH/NCCD Lung Cancer Dataset, which is available in the Mendeley Data repository at https://data.mendeley.com/datasets/bhmdr45bh2/2.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

Acknowledgement

References

[1]   P. Monkam, S. Qi, H. Ma, W. Gao, Y. Yao, W.Q.-I. Access, undefined 2019, Detection and classification of pulmonary nodules using convolutional neural networks: a survey, Ieeexplore.Ieee.OrgP Monkam, S Qi, H Ma, W Gao, Y Yao, W QianIeee Access, 2019•ieeexplore.Ieee.Org (n.d.). https://ieeexplore.ieee.org/abstract/document/8736217/ (accessed June 13, 2025).

[2]   V. Kumar, B. Bakariya, Classification of malignant lung cancer using deep learning, J. Med. Eng. Technol. 45 (2021) 85–93. https://doi.org/10.1080/03091902.2020.1853837.

[3]   R. Fontana, D. Sanderson, … W.T.-A.R. of, undefined 1984, Early lung cancer detection: results of the initial (prevalence) radiologic and cytologic screening in the Mayo Clinic study, Atsjournals.OrgRS Fontana, DR Sanderson, WF Taylor, LB Woolner, WE Miller, JR Muhm, MA UhlenhoppAmerican Review of Respiratory Disease, 1984•atsjournals.Org (n.d.). https://www.atsjournals.org/doi/abs/10.1164/arrd.1984.130.4.561 (accessed June 13, 2025).

[4]   S.B. Knight, P.A. Crosbie, H. Balata, J. Chudziak, T. Hussell, C. Dive, Progress and prospects of early detection in lung cancer, Royalsocietypublishing.OrgS Blandin Knight, PA Crosbie, H Balata, J Chudziak, T Hussell, C DiveOpen Biology, 2017•royalsocietypublishing.Org 7 (2017). https://doi.org/10.1098/rsob.170070.

[5]   R.L. Siegel Mph, A.N. Giaquinto, | Ahmedin, J. Dvm, R.L. Siegel, Cancer statistics, 2024, (2024). https://doi.org/10.3322/caac.21820.

[6]   T. Inage, T. Nakajima, … I.Y.-C. in chest, undefined 2018, Early lung cancer detection, Chestmed.Theclinics.Com (n.d.). https://www.chestmed.theclinics.com/article/S0272-5231(17)30107-7/abstract (accessed June 13, 2025).

[7]   S. Deepajothi, R. Balaji, T. Madhuvanthi, P. Priakanth, T. Daniya, S. Velliangiri, Detection and Stage Classification of UNet Segmented Lung Nodules Using CNN, 2022 5th International Conference on Multimedia, Signal Processing and Communication Technologies, IMPACT 2022 (2022). https://doi.org/10.1109/IMPACT55510.2022.10029182.

[8]   N. Mohanapriya, B. Kalaavathi, T.S. Kuamr, Lung Tumor Classification and Detection from CT Scan Images using Deep Convolutional Neural Networks (DCNN), Proceedings of 2019 International Conference on Computational Intelligence and Knowledge Economy, ICCIKE 2019 (2019) 800–805. https://doi.org/10.1109/ICCIKE47802.2019.9004247.

[9]   G. Li, W. Zhou, W. Chen, F. Sun, Y. Fu, … F.G.-I., undefined 2020, Study on the detection of pulmonary nodules in CT images based on deep learning, Ieeexplore.Ieee.OrgG Li, W Zhou, W Chen, F Sun, Y Fu, F Gong, H ZhangIEEE Access, 2020•ieeexplore.Ieee.Org (n.d.). https://ieeexplore.ieee.org/abstract/document/9050800/ (accessed June 13, 2025).

[10]   O. Ozdemir, R. Russell, A.B.-I. transactions on, undefined 2019, A 3D probabilistic deep learning system for detection and diagnosis of lung cancer using low-dose CT scans, Ieeexplore.Ieee.OrgO Ozdemir, RL Russell, AA BerlinIEEE Transactions on Medical Imaging, 2019•ieeexplore.Ieee.Org (n.d.). https://ieeexplore.ieee.org/abstract/document/8886352/ (accessed June 13, 2025).

[11]   W. Salama, A. Shokry, M.A.-M.T. and Applications, undefined 2022, A generalized framework for lung Cancer classification based on deep generative models, SpringerWM Salama, A Shokry, MH AlyMultimedia Tools and Applications, 2022•Springer 81 (2022) 32705–32722. https://doi.org/10.1007/s11042-022-13005-9.

[12]   Y. Cakmak, A. Maman, Deep Learning for Early Diagnosis of Lung Cancer, Computational Systems and Artificial Intelligence 1 (2025) 20–25. https://doi.org/10.69882/ADBA.CSAI.2025074.

[13]   Y. Cakmak, Machine Learning Approaches for Enhanced Diagnosis of Hematological Disorders, Computational Systems and Artificial Intelligence 1 (2025) 8–14. https://doi.org/10.69882/ADBA.CSAI.2025072.

[14]   Y. Cakmak, I. Pacal, AI-Driven Classification of Anemia and Blood Disorders Using Machine Learning Models, Computers and Electronics in Medicine 2 (2025) 43–52. https://doi.org/10.69882/ADBA.CEM.2025073.

[15]   Pacal Ishak, Cakmak Yigitcan, DIAGNOSTIC ANALYSIS OF VARIOUS CANCER TYPES WITH ARTIFICIAL INTELLIGENCE, 2025. www.duvaryayinlari.com.

[16]   I. Kayadibi, G.G.-I.J. for Multiscale, undefined 2023, An early retinal disease diagnosis system using OCT images via CNN-based stacking ensemble learning, Dl.Begellhouse.ComI Kayadibi, GE GüraksınInternational Journal for Multiscale Computational Engineering, 2023•dl.Begellhouse.Com (n.d.). https://www.dl.begellhouse.com/journals/61fd1b191cf7e96f,5a368e170ddc2b7e,69291d0660911da 2.html (accessed June 13, 2025).

[17]   İ. Kayadibi, G. Güraksın, U.K.-E.S. with Applications, undefined 2023, A Hybrid R-FTCNN based on principal component analysis for retinal disease detection from OCT images, Elsevierİ Kayadibi, GE Güraksın, U KöseExpert Systems with Applications, 2023•Elsevier (n.d.). https://www.sciencedirect.com/science/article/pii/S0957417423011193 (accessed June 13, 2025).

[18]   D. Coşkun, D. Karaboğa, A. Baştürk, B. Akay, Ö.U. Nalbantoğlu, S. Doğan, İ. Paçal, M.A. Karagöz, A comparative study of YOLO models and a transformer-based YOLOv5 model for mass detection in mammograms, Turkish Journal of Electrical Engineering and Computer Sciences 31 (2023) 1294–1313. https://doi.org/10.55730/1300-0632.4048.

[19]   I. Pacal, O. Attallah, InceptionNeXt-Transformer: A novel multi-scale deep feature learning architecture for multimodal breast cancer diagnosis, Biomed. Signal Process. Control 110 (2025) 108116. https://doi.org/10.1016/J.BSPC.2025.108116.

[20]   İ. Pacal, Deep Learning Approaches for Classification of Breast Cancer in Ultrasound (US) Images, Journal of the Institute of Science and Technology 12 (2022) 1917–1927. https://doi.org/10.21597/JIST.1183679.

[21]   G. Işık, İ. Paçal, Few-shot classification of ultrasound breast cancer images using meta-learning algorithms, Neural Comput. Appl. 36 (2024) 12047–12059. https://doi.org/10.1007/S00521-024-09767-Y/TABLES/7.

[22]   Y. Cakmak, I. Pacal, Enhancing Breast Cancer Diagnosis: A Comparative Evaluation of Machine Learning Algorithms Using the Wisconsin Dataset, Journal of Operations Intelligence 3 (2025) 175–196. https://doi.org/10.31181/JOPI31202539.

[23]   Y. Cakmak, S. Safak, M.A. Bayram, I. Pacal, Comprehensive Evaluation of Machine Learning and ANN Models for Breast Cancer Detection, Computer and Decision Making: An International Journal 1 (2024) 84–102. https://doi.org/10.59543/COMDEM.V1I.10349.

[24]   Y. Cakmak, J. Zeynalov, A Comparative Analysis of Convolutional Neural Network Architectures for Breast Cancer Classification from Mammograms, Artificial Intelligence in Applied Sciences 1 (2025) 28–34. https://doi.org/10.69882/ADBA.AI.2025075.

[25]   Y. Cakmak, N. Pacal, Deep Learning for Automated Breast Cancer Detection in Ultrasound: A Comparative Study of Four CNN Architectures, Artificial Intelligence in Applied Sciences 1 (2025) 13–19. https://doi.org/10.69882/ADBA.AI.2025073.

[26]   A. Karaman, I. Pacal, A. Basturk, B. Akay, U. Nalbantoglu, S. Coskun, O. Sahin, D. Karaboga, Robust real-time polyp detection system design based on YOLO algorithms by optimizing activation functions and hyper-parameters with artificial bee colony (ABC), Expert Syst. Appl. 221 (2023) 119741. https://doi.org/10.1016/J.ESWA.2023.119741.

[27]   A. Karaman, D. Karaboga, I. Pacal, B. Akay, A. Basturk, U. Nalbantoglu, S. Coskun, O. Sahin, Hyper-parameter optimization of deep learning architectures using artificial bee colony (ABC) algorithm for high performance real-time automatic colorectal cancer (CRC) polyp detection, Applied Intelligence 53 (2023) 15603–15620. https://doi.org/10.1007/S10489-022-04299-1/TABLES/6.

[28]   I. Pacal, MaxCerVixT: A novel lightweight vision transformer-based Approach for precise cervical cancer detection, Knowl. Based. Syst. 289 (2024) 111482. https://doi.org/10.1016/J.KNOSYS.2024.111482.

[29]   I. Pacal, S. Kılıcarslan, Deep learning-based approaches for robust classification of cervical cancer, Neural Comput. Appl. 35 (2023) 18813–18828. https://doi.org/10.1007/S00521-023-08757-W/FIGURES/8.

[30]   S. Ince, I. Kunduracioglu, A. Algarni, B. Bayram, I. Pacal, Deep learning for cerebral vascular occlusion segmentation: A novel ConvNeXtV2 and GRN-integrated U-Net framework for diffusion-weighted imaging, Neuroscience 574 (2025) 42–53. https://doi.org/10.1016/J.NEUROSCIENCE.2025.04.010.

[31]   I. Pacal, A novel Swin transformer approach utilizing residual multi-layer perceptron for diagnosing brain tumors in MRI images, International Journal of Machine Learning and Cybernetics 15 (2024) 3579–3597. https://doi.org/10.1007/S13042-024-02110-W/TABLES/5.

[32]   I. Pacal, O. Akhan, R.T. Deveci, M. Deveci, NeXtBrain: Combining local and global feature learning for brain tumor classification, Brain Res. 1863 (2025) 149762. https://doi.org/10.1016/J.BRAINRES.2025.149762.

[33]   I. Pacal, O. Celik, B. Bayram, A. Cunha, Enhancing EfficientNetv2 with global and efficient channel attention mechanisms for accurate MRI-Based brain tumor classification, Cluster Comput. 27 (2024) 11187–11212. https://doi.org/10.1007/S10586-024-04532-1/TABLES/4.

[34] J. Debnath, A.S. Uddin Khondakar Pranta, A. Hossain, A. Sakib, H. Rahman, R. Haque, M.R. Ahmed, A.W. Reza, S.M.M.R. Swapno, A. Appaji, LMVT: A hybrid vision transformer with attention mechanisms for efficient and explainable lung cancer diagnosis, Inform. Med. Unlocked 57 (2025) 101669. https://doi.org/10.1016/J.IMU.2025.101669.

[35] A.J. Yagappan, H. Karuppiah, M. Muthusamy, S.K. Kannaiah, Optimizing lung cancer classification using transformer and Gooseneck Barnacle Optimization, Expert Syst. Appl. 282 (2025) 127413. https://doi.org/10.1016/J.ESWA.2025.127413.

[36] D. Mannepalli, T. Kuan Tak, S. Bala Krishnan, V. Sreenivas, GSC-DVIT: A vision transformer based deep learning model for lung cancer classification in CT images, Biomed. Signal Process. Control 103 (2025) 107371. https://doi.org/10.1016/J.BSPC.2024.107371.

[37] J. Tang, X. Chen, L. Fan, Z. Zhu, C. Huang, LN-DETR: An efficient Transformer architecture for lung nodule detection with multi-scale feature fusion, Neurocomputing 633 (2025) 129827. https://doi.org/10.1016/J.NEUCOM.2025.129827.

[38] F. Cui, Y. Li, H. Luo, C. Zhang, H. Du, SF2T:Leveraging Swin Transformer and Two-stream networks for lung nodule detection, Biomed. Signal Process. Control 95 (2024) 106389. https://doi.org/10.1016/J.BSPC.2024.106389.

[39] W. Li, X. Zou, J. Zhang, M. Hu, G. Chen, S. Su, Predicting lung cancer bone metastasis using CT and pathological imaging with a Swin Transformer model, J. Bone Oncol. 52 (2025) 100681. https://doi.org/10.1016/J.JBO.2025.100681.

[40] L.F. Talib, J. Amin, M. Sharif, M. Raza, Transformer-based semantic segmentation and CNN network for detection of histopathological lung cancer, Biomed. Signal Process. Control 92 (2024) 106106. https://doi.org/10.1016/J.BSPC.2024.106106.

[41] H.F. Kareem, M.S. AL-Huseiny, F. Y. Mohsen, E. A. Khalil, Z. S. Hassan, Evaluation of SVM performance in the detection of lung cancer in marked CT scan dataset, Indonesian Journal of Electrical Engineering and Computer Science 21 (2021) 1731. https://doi.org/10.11591/IJEECS.V21.I3.PP1731-1738.).

[42] The IQ-OTH/NCCD lung cancer dataset, (n.d.). https://www.kaggle.com/datasets/hamdallak/the-iqothnccd-lung-cancer-dataset (accessed August 3, 2025).

[43] Z. Wang, P. Wang, K. Liu, P. Wang, Y. Fu, C.-T. Lu, C.C. Aggarwal, J. Pei, Y. Zhou, A Comprehensive Survey on Data Augmentation, (2024). https://arxiv.org/abs/2405.09591v3 (accessed May 28, 2025).

[44] A. Mumuni, F. Mumuni, N.K. Gerrar, A Survey of Synthetic Data Augmentation Methods in Machine Vision, Machine Intelligence Research 2024 21:5 21 (2024) 831–869. https://doi.org/10.1007/S11633-022-1411-7.

[45] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby, AN IMAGE IS WORTH 16X16 WORDS:

TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE, (n.d.). https://github.com/ (accessed August 3, 2025).

[46]   H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, H. Jégou, F. Ai, Training data-efficient image transformers & distillation through attention, (n.d.).

[47]   Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo, Swin Transformer: Hierarchical Vision Transformer using Shifted Windows, (n.d.). https://github. (accessed July 1, 2025).

[48]   H. Bao, L. Dong, S. Piao, F. Wei, BEIT: BERT Pre-Training of Image Transformers, (n.d.). https://aka.ms/beit (accessed July 1, 2025).