scientific reports

OPEN

# Enhancing lung cancer detection through integrated deep learning and transformer models

Revathi Durgam[1], Bharathi Panduri[2], V. Balaji[3], Adil O. Khadidos[4], Alaa O. Khadidos[5,9] & Shitharth Selvarajan[6,7,8]✉

**Lung cancer has been stated as one of the prevalent killers of cancer up to this present time and this clearly underlines the rationale for early diagnosis to enhance life expectancy of patients afflicted with the condition. The reasons behind the usage of the transformer and deep learning classifiers for the detection of lung cancer include accuracy, robustness along with the capability to handle and evaluate large data sets and much more. Such models can be more complex and can help to utilize multiple modalities of data to give extensive information that will be critical in ascertaining the right diagnosis at the right time. However, the existing works encounter several limitations including reliance on large annotated data, overfitting, high computation complexity, and interpretability. Third, the issue of the stability of these models' performance when applied to actual clinical datasets is still an open question; this is an even bigger issue that will greatly reduce the actual utilization of these models in clinical practice. To tackle these, we develop a novel Cancer Nexus Synergy (CanNS), which applies of A. Swin-Transformer UNet (SwiNet) Model for segmentation, Xception-LSTM GAN (XLG) CancerNet for classification, and Devilish Levy Optimization (DevLO) for fine-tuning parameters. This paper breaks new ground in that the presented elements are incorporated in a manner that co-operatively elevates the diagnostic capabilities while at the same time being computationally light and resilient. These are SwiNet for segmented analysis, XLG CancerNet for precise classification of the cases, and DevLO that optimizes the parameters of the lung cancer detection system, making the system more sensible and efficient. The performance outcomes indicate that the CanNS framework enhances the detection's accuracy, sensitivity, and specificity compared to the previous approaches.**

**Keywords** Lung Cancer, Disease detection, Transformer models, Deep learning, Optimization, Segmentation, And classification

Lung nodules are small, spherical growths in the lung that appear as areas of increased density on CT images[1]. These nodules may be caused by conditions such as smoking, which exposes the lungs to extremely dangerous chemicals, or from long-term inhalation of fume-like poisons like dust, chemicals, or asbestos[2]. Early detection and classification of lung nodules in diagnosing lung diseases are very important because early intervention at the right time can be done. Accurately identifying these nodules assists health experts in distinguishing between benign growths and possible malignancies. This can drastically alter a patient's prognosis and the line of treatment. Hence, early detection enables better management of the lung condition and possibly reduces the burden of the disease, thereby improving survival rates. Lung cancer is the type of cancer originating in the lungs and represents the foremost cause of cancer-related death worldwide. It is usually classified into two broad categories: small cell lung cancer and non-small cell lung cancer[3], with quite differing characteristics, modes of treatment, and prognosis. It may be genetic, environmental exposure, or a result of an individual's lifestyle choices.

[1]Department of Data Science, AVN Institute of Engineering and Technology, Hyderabad, India. [2]Department of Information Technology, Gokaraju Rangaraju Institute of Engineering & Technology, Hyderabad, India. [3]Department of CSE-AIML, Vardhaman College of Engineering, Hyderabad, India. [4]Department of Information Technology, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia. [5]Department of Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia. [6]Department of Computer Science and Engineering, Kebri Dehar University, Kebri Dehar, Ethiopia. [7]Centre for Research Impact & Outcome, Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, Rajpura, 140401, India. [8]Department of Computer Science and Engineering, Chennai Institute of Technology, Chennai, India. [9] Center of Research Excellence in Artificial Intelligence and Data Science, King Abdulaziz University, Jeddah, Saudi Arabia. ✉email: shitharths@kdu.edu.et

Smoking is the biggest risk factor. Other factors include secondhand smoke, exposure to radon gas, asbestos, and other hazardous environmental elements. Early detection of the disease is an extremely critical factor in lung cancer[4]. Most cases do not present symptoms of lung cancer in the early stages, so it is indeed very hard to detect it without regular screenings in the form of low-dose CT scans. In places where lung cancer is detected at an early stage, the chances of successful treatment and survival rate become considerably high. The early detection opens up more options for treatment, such as surgical removal of tumors, targeted therapy, radiation, and chemotherapy. This will also help in preventing the spread of cancer to other parts of the body, which would complicate treatment and lower survival rates. Early diagnosis and treatment of lung cancer will help one avoid symptoms, improve quality of life, and also reduce the economic burden of advanced cancer care. Advanced imaging techniques and predictive models give a great deal of help in terms of identification and tracking of nodules. In this way, health providers will be well-equipped to make a timely decision about treatment against lung cancer. Through this proactive approach, it results in improved patient outcome. Healthier patients imply more efficient healthcare systems that reduce the need for extensive late-stage, and very expensive treatments.

Artificial Intelligence (AI)[5,6] based lung cancer detection and diagnosis, becoming ever more precise and efficient. It analyzes medical images, such as computed tomography scans, with a high degree of accuracy for the presence of cancer, much better than old traditional methods of diagnosis. As such, technological development has been very critical to the early detection and diagnosis of lung cancer, since it bears greatly on outcome variation among the patients[7,8]. Key among the advantages that AI can bring to the detection of lung cancer is the processing of huge amounts of data quickly and perfectly by algorithms. Such algorithms can detect slight patterns and anomalies in medical imaging that went unnoticed by the experts' eyes, thereby increasing sensitivity and specificity in lung cancer screening. This will give increased detection capability in the identification of nodules in the lungs to a much earlier stage, wherein timely initiation of treatment may be the ones that save lives. On the other hand, AI-based systems[9] are also associated with known consistency and lack of bias, hence reducing variability and probable errors due to human interpretation. This is very important in ensuring that every patient undergoes a complete and accurate assessment, totally independent of the workload or experience of the radiologist. If need be, AI can help detect changes in lung nodules over time, which is a very essential feature of progression or regression of illnesses likely to formulate an effective mode of treatment and monitoring. AI integration in the lung cancer detection workflow will grossly automate the diagnostic procedure, improving efficiency while taking some significant workload off healthcare professionals' shoulders. AI is able to empower the radiologist and the oncologist to spend more time with the patient and concentrate on the more demanding decision-making process by automating routine tasks and providing rapid analysis possibilities[10,11]. This will not only help enhance the quality of health delivery services but will also, in some fashion, cope sustainably with growing diagnostic imaging volumes.

The class of AI models that the transformer models are—the first of their kind to be developed for processing natural language—have recently shown great promise in medical imaging, say in lung cancer detection. One of their mechanisms is self-attention, which confers on these models the ability to compute the importance of various elements in an input sequence relative to each other[12,13]. This gives them the capability to pick up complex patterns and relationships from data. In the beginning, transformers were used in NLP; they are a deep neural network structure mainly using the self-attention mechanism to extract inherent characteristics from text. Because of the versatility of transformers for representation and learning, researchers started to examine the ways to apply transformers for computer vision tasks. Text data does not pose such difficulties; however, processing image data introduces extra problems[14] due to the massive amount of data, high noise level, and multiple modalities that make it difficult to apply transformers efficiently. Nevertheless, transformers have recently attained substantial advancements in the domain of computer vision and perform exceedingly well on natural images. This integration has been utilized in very few fields particularly in multi-modal medical image fusion in the medical image analysis field. Multi-modal images[15] are extensively employed in the medical image analysis for the better disease diagnosis and lesion extraction because it integrates different modes to provide more detailed information about the patient's health. Existing approaches to medical image multi-modal fusion using deep learning can be categorized into three primary types: In the context of multi-sensor data fusion, the first's type includes input-level fusion, feature-level fusion, and decision-level fusion. Input level fusion entails the joining of different modalities and passing them into the neural network at the input level. This approach covers the attempt to merge raw data, which may come from different sources, into one common input. Feature-level fusion integrates the features of the different modalities in the network. This method leverages on the characteristics that accompany each modality since the features are combined, yielding improved and more reliable representations[16]. The final fusion process integrates the output or decision generated by the different networks that are dealing with different modalities. This strategy is possible because the final decision is made after the analysis of each modality, and the results of all the analyses performed for the specific modality are summarized.

Some of the major benefits associated with transformer models in lung cancer detection are very many and significant, particularly in the line of improving diagnostic accuracy and efficiency. How transformers process large amounts of high-dimensional data makes them rather fitting for the analysis of these complex medical images to identify and classify pulmonary nodules correctly, hence giving more reliable diagnoses[17,18]. Their self-attention mechanism is capable of perceiving the entire image context, therefore enhancing their ability to detect very fine and early signs of cancer, which might be context dependent and important to make distinctions between benign and malignant nodules. Another critical advantage of the transformer models is scalability, for which it can be further scaled up by making use of larger datasets and more computational resources to guarantee continuous improvement in diagnostic accuracy with an increased quantity of medical data. Furthermore, transformers actually leveraged transfer learning: that is, pre-trained models on extensive datasets that can be fine-tuned for specific tasks like the detection of lung cancer, thereby lessening the requirement of large labeled datasets and speeding up the pace through which accurate diagnostic models can be developed[19]. They also excel at multimodal integration—fuse images from CT into patient records, genomic data, and all other relevant information to provide a complete overview of the condition of a patient, therefore better diagnosis and treatment decisions. Besides, this will also lead to workload reduction from healthcare professionals because the transformer models will automate the image analysis process, and these professionals can focus on other more critical aspects of patient care[20]. This will ensure faster turn-around times for diagnostic results and hence enhanced overall healthcare efficiency and outcome.

While deep learning and earlier models of transformers have undoubtedly pushed the progress in lung cancer detection, clinical applications have had formidable challenges. Most fundamentally, they require large, scrupulously annotated datasets that train models for the recognition of patterns indicative of lung cancer on medical images such as CT scans[14,21]. Acquisition of these datasets involves much effort and is time-consuming, sometimes requiring an expert radiologist for proper annotation of the images. Furthermore, deep learning models are prone to overfitting, such that the models might remember certain nuances of the training dataset but at the same time fail well on another patient dataset, hence causing reduced diagnostic accuracy in real-world clinical settings. This can be a severe challenge in the case of lung cancer detection, since there is wide variability in the imaging conditions and patient demographics that influence the appearance of tumors and nodules in CTs[22,23]. Although transformer models have good performance in learning complex features from images, they are adding extra computational loads. This architecture, especially the mechanism of self-attention that confers on them the contextual understanding across images, involves a tremendous number of computational resources in training and inference. That is what produces computational intensity, hence limiting the practical deployment of these models, especially in healthcare facilities with constrained hardware capabilities and not well served with high-performance computing resources. These limitations call manifold for the dire need for innovative approaches in lung cancer detection and thus, at the very outset, the proposed for medical image analysis. The main objectives of the proposed work are given below:

- To design and implement a CanNS framework that integrates SwiNet segmentation with XLG CancerNet classification techniques for the accurate detection of lung cancers.
- To develop and optimize image segmentation algorithms using SwinNet to support highly accurate identification in the location and delineation of lung nodules/cancerous areas on CT images.
- The further classification of the segmented lung regions was supported by an XLG cancer network deep learning architecture with high accuracy, demarcating benign from malignant nodules.
- Seek for the best hyper-parameters of the model using Devilish Levy Optimization techniques for optimum learning rates and ensuring high-performance models to diagnose lung cancer effectively.

The other portions of this paper are segregated into the following sections: Sect. "Related works" gives a background on the current state of deep learning and transformer models applied for lung CT cancer diseases detection, revisiting previous methodologies underlining their strengths and limitations and mentioning new and recent developments. This will be of importance in arriving at an overall understanding of the state-of-the-art techniques that can lend support to the evaluation and contextualizing of the proposed CanNS framework. Section "Proposed methodology" will deal with the CANNS model, specifying novelty behind the approach. This would pertain to objectives to be met in the framework, details of the workflow from image preprocessing to SwiNet segmentation, going all the way to XLG CancerNet classification, and finally explaining the architectural designs and the mathematical models involved at each stage. Section "Results and discussion" deals with the modeling of empirical results and comparative analyses of the performance of the CanNS model against existing benchmarks for accuracy, sensitivity, sensitivity, and all other relevant metrics. Section "Conclusion" concludes synthesis by specifying the practical implications for clinical settings of the proposed CanNS framework and the avenues for further research and improvement to increase its efficacy and applicability in lung-CT-based cancer detection.

## Related works

This section reviews recent advances in deep learning and transformer models tailored toward the detection of lung cancer diseases, exploring how these technologies have evolved within medical imaging, with emphasis on bettering diagnostic accuracy and efficiency. This discussion presents a myriad of methodologies, spanning from early CNNs[24,25] to more recent transformer-based architectures, pointing out the strengths and limitations for each approach applied in the analysis of computed tomography scans of lung images. This review should give an overview of the existing approaches based on one view of the literature and therefore set the stage for the introduction of the proposed CanNS framework.

Quasar, et al.[26] makes use of the Chest CT-Scan Images dataset for comparing different approaches with regard to lung cancer diagnosis. Much research conducted in the recent past has been done by ensemble methods, deep learning techniques, or even hybrid models for the sake of overcoming the odds of accurate diagnosis. Of these, ensemble methods have been particularly very instrumental in integrating different classifiers since they borrow from the strengths of the different models to offset their weaknesses, hence improving general performance. One of the expected disadvantages of ensemble methods is that added complexity in model interpretability and implementation can be very resource-demanding in terms of computational resources and expertise needed to optimize these models. Despite the effort, results from this study prove the proposed ensemble approach outperforms single-model techniques and any other existing ensemble methods in diagnosis accuracy and reliability. Shah, et al.[27] reviewed state-of-the-art methodologies for automated lung cancer detection and classification from radiological data, with emphasis on how machine learning and deep learning techniques could play a key role in providing accurate, trustworthy, and rapid assessments on disease progression. The present study has thus underlined how advanced algorithms increase the accuracy and speed of diagnosis significantly in aiding clinical decision-making. Most of these methods require large, annotated training datasets; this kind of data is usually hard to come by. Moreover, the models generally require huge computational powers and mostly have problems like overfitting, where it works well only on training data and fails generalization to new unseen data. Moreover, the deep learning models per se can be very complex and hence opaque to interpretability itself, making it difficult for clinicians to use them in decision making. Notwithstanding these limitations, the potential for combining machine learning and deep learning in the diagnosis of lung cancer remains bright, with very encouraging improvements in diagnostic accuracy and efficiency. Thanoon, et al.[28] presented a deep review of the developed techniques of deep learning for lung cancer screening and diagnosis. In this review, an overview of deep learning techniques is done where, with reference to its application in lung cancer, emphasis is placed on suggested DL methods and reviewed unique features by methodology, thereby focusing on two primary methodologies: that of classification and segmentation. Basically, the classification methodology is related to the presence of cancer, while segmentation delineates the cancerous regions in the lung images. The review provides advantages of the current DL models with respect to large dataset processing capabilities and its high accuracy for the detection of lung cancer. It also discusses shortcomings like the requirement of large amounts of annotated datasets and the risk of overfitting and high computational resources. This review recognizes the power of DL as a very promising tool toward better diagnosis of lung cancer while pointing out challenges to its general clinical application.

Siddiqui, et al.[29] proposed a new approach to classify the images of CT scans of lungs with the help of Gabor filters in combination with an enhanced Deep Belief Network in conjunction with several classification methods for betterment in processing time and accuracy. E-DBN uses cascading two Restricted Boltzmann Machines: Gaussian-Bernoulli and Bernoulli-Bernoulli RBMs. In this work, SVM was applied and it showed good efficiency with the best performance parameters among the applied ones. By integrating SVM and E-DBN into the system, the proposed model would improve its classifications in lung CT images for better accuracy and faster processing times. This model can exploit the strengths of Gabor filters, E-DBN, and SVM to help resolve some of the common issues, particularly in the classification for lung cancer diagnosis. However, it cannot circumvent the dependence on the quality of the initial feature extraction and possible complexity in joining multiple techniques, thus challenging practical reliability and scalability. Nigudgi, et al.[30] proposed a transfer learning-based methodology for the classification of lung CT images and cancer detection. During the process, the input images will be fed to a pre-trained hybrid model that is a combination of AlexNet, VGG, and GoogleNet portraying features from the images. After feature extraction of these images using this hybrid model, classification will be fed into a multi-class SVM. Therefore, the methodology would be to train the model on a dataset and test it on real-time images to prove its effectiveness. This shall, in essence, encapsulate the power of the deep models learnt by the pre-trained models in relation to the SVM; however, their performance heavily relies upon the pre-trained models and the feature extraction process. This hybrid model could therefore involve high computational cost, turning out to be hard to apply in real-time applications due to latency associated with added complexity. Lanjewar, et al.[31] proposed a new DL-based method by tuning DenseNet201, which added new layers to the original DenseNet framework in identifying the lung cancer disease. Then, they proposed two methods of feature selection to extract the best features from DenseNet201 and applied the same in various other machine learning classifiers for final classification. While this offers an efficient way to integrate deep learning for feature extraction with traditional machine learning classifiers, increasing accuracy, it does not come without a few disadvantages. Additional layers atop DenseNet will raise computational complexity and further extend an already long training time. On the other hand, relying on feature selection methods might increase variability in performances with respect to qualities and relevance of features to be selected. This will also result in excess and heavy work with different ML classifiers whose performance in various data sets is far from optimal.

AR Bushara, et al.[32] proposed a deep learning framework of novel design, called LCD-CapsNet, which encapsulates a convolutional neural network and a capsule neural network to leverage the capabilities of major networks in the application of lung cancer detection and classification using computed tomography images. The framework is suggested to learn with minimized requirements of data and acquire spatial invariance attributes related to the enhanced accuracy of lung cancer diagnosis. The main objectives of the proposed method are to develop algorithms for classifying and analyzing the images in a dataset to identify whether the patient has or is prone to lung cancer. Yet, it has the key problems like the requirement of more computational resources and more time taken to train the model. Also, the effectiveness is heavily dependent on the quality and diversity of a training dataset. In that case, generalizability to different patient populations and image conditions will be very poor. Moreover, performance could be sensitive to anything from hyperparameter tuning to the architectural design choices of the model, thus requiring a great deal of experimentation and optimization. Dunn, et al.[33] used the publicly available trained incremental multiple resolution residual network deep learning segmentation model to auto-segment the CT images. The multi-resolution pathways and residual connections in the model iMRRN allow the latter to ensure accuracy and efficiency in its segmentation. However, the approach has its disadvantages. This might reduce their flexibility to adapt to new and different datasets and may then contribute less to generalization across imaging conditions or patient populations. Moreover, such a design is computationally very demanding and complex for any model; hence, it is hardly deployable in resource-constrained settings.

Although deep learning and transformer models[14] have considerably improved in disease detection for lung cancer, there are a number of research gaps that exist and are impeding the optimal diagnostic performance. Of these, one of the most important is their dependency on large, annotated datasets for training, which, in the domain of medical imaging, is hard to obtain because each image needs to be annotated by experts. This most commonly leads to models that cannot generalize to other populations of patients and variations in acquisition conditions. Deep learning models[34] also notoriously overfit, where performance is good on the training data but poor on unseen data, limiting real-world clinical applicability. Another key gap concerns the strong computational cost related to the training and deployment of transformer models. These representations require, however, enormous hardware resources, making them less accessible for too many institutions in the healthcare sector[16], let al.one those in resource-limited settings. Moreover, while transformer models show some promise in large-scale dataset processing in capturing complex patterns, their application in medical imaging—mostly related to multi-modal integration—is still in its nascent stages of development. Further research is required on the integration of these different kinds of medical data, from CT images through patient records to genomic data, in a manner that leads to better diagnosis and health outcomes[35]. Influential in this are the facts that most of the deep learning and transformer models are poorly interpretable, which is necessary to gain the trust of clinicians by ensuring AI-driven decisions are transparent and understandable. Basically, these models are black boxes, and this has given rise to many challenges in explaining their decision-making processes—this being the main obstacle to their clinical application. One of the most important of these relates to the strength of these models in handling real-world clinical data, which are generally noisy and incomplete[36]. There is a need therefore to develop models that can handle such data robustly but still deliver accurate and reliable diagnoses. This would require addressing these research gaps to move the field of AI-driven lung cancer detection[37,38] forward and to facilitate the translation of these technologies into more effective and widely adopted clinical practices.

## Proposed methodology

This section presents a coherent narrative of the rationale for the proposed Cancer Nexus Synergy (CanNS) Framework for diagnosing the cancer diseases. This smart and novel framework, developed in the proposed work, incorporates three distinct and unique methodologies such as SwiNet, XLG, and DevLO. The CanNS Framework is proposed in order to incorporate the state-of-art methods in the specific fields of image segmentation, feature extraction, temporal pattern analysis and optimization providing the medical images based cancer diagnostic system to be efficient and coherent. The flow of the proposed CanNS model is shown in Fig. 1. The first operation involves SwiNet model that is currently recognized as the leading model in image segmentation. SwiNet is based on the Swin-Transformer since it is suitable for the given task as the model pays equal attention to the micro and macro-levels of an image. SwiNet split the input medical images into several non-overlapping patches, and feeds the latter through the multiple transformer blocks during processing. The proposed SwiNet model performs better in capturing both the local and global feature, enhanced the degree of segmentation and capable to handle high resolution image effectively. Besides, this work helps to advance the creation of improved models for cancer detection and foster new advances in utilizing Vision Transformers for different tasks in medical imaging. In the specific domain of cancer detection, SwiNet offer a total scheme to solve the task of segmenting tumor regions in medical images. The cleaning stage guarantees that it is possible to work with data that is clean and is standard for analysis. The structure of Swin Transformer blocks with a hierarchical mechanism to extract features at multiple scales is considered advantageous when detecting tumors of different sizes and shapes. It can also be seen in the ideas of the local-global attention mechanisms that SwiNet has better parameterization characteristics in modelling the structures and relations of medical images and further enhances the correct degree of tumor delineation. The patch merging and the expanding operations improve the computational speed, by making the processing of large number of high resolution medical images without consuming high amount of resources possible and practical. UNet architecture has the skip connections to maintain high-resolution features that enable the final segmentation map to have the right details for diagnosis. Overall, the proposed SwiNet model is performed well in the tasks of cancer image segmentation based on the latest Vision Transformers technique compared to the previous models. It improves cancer detection's precision and advances the state of the art of Vision Transformers for medical image analysis, there is evidence for the applicability of vision transformers for difficult medical imaging problems.

This multi-scale approach ensures that the segmentation is properly done, and the regions of interest true and only the possible regions will be isolated that can contain cancerous tissues. It is important as it is the basis for the subsequent phases that a thorough proper spotting of possible malignant areas appropriately is done. After that, this stage of segmentation is completed to finally use the Xception-LSTM GAN (XLG) CancerNet for classification. XLG CancerNet integrates three powerful components: Particularly, there are Xception Net, Deep LSTM, and GAN. The Xception Net does depthwise separable convolution; it applies efficient methods to extract final features from a separated image. These features limit the detail created in medical imaging which is useful in differentiating benign and malignant cells in the tissues. These derived features are then reshaped and passed into a Deep LSTM network that is useful in identifying time dependent patterns. This comes in conveniently after change in the nature of medical image where sequel of the diseases can most times be inferred from sequential data. They are complemented with GANs that make precise fake images which, in their turn, enlarge the training set and stabilize and improve the model. Such classification proves to be extremely stable in relation to variations of inputs and at the same time learns the manner of differentiation of cancer tissues from non-cancer ones with a level of accuracy that is comparable to that of a human pathologist.

The devilish Levy optimization supports the search for further performance optimization within CanNS. Of all the critical hyper-parameters, learning rate has immense effects not only on the behavior of model training dynamics but also on the speed of convergence. DevLO is an algorithm inspired by Tasmanian Devil Optimization with Levy flights that enables a dynamic and adaptive method of fine-tuning learning rates. The technique wonderfully mimics the natural foraging behavior of Tasmanian devils by intelligently exploring this space and creating a trade-off between the exploitation and exploration of new possibilities. DevLO ensures that the learning rate is optimized so XLG CancerNet converges to its optimal performance, hence leading to more accurate and reliable diagnosis for cancer patients. The CanNS Framework: A broad, singular approach to the diagnosis of cancer from medical images. Because these are the newest methodologies vividly integrated with each other, diagnostic accuracy, robustness, and efficiency are very high. Precise segmentation by SwiNet, robust

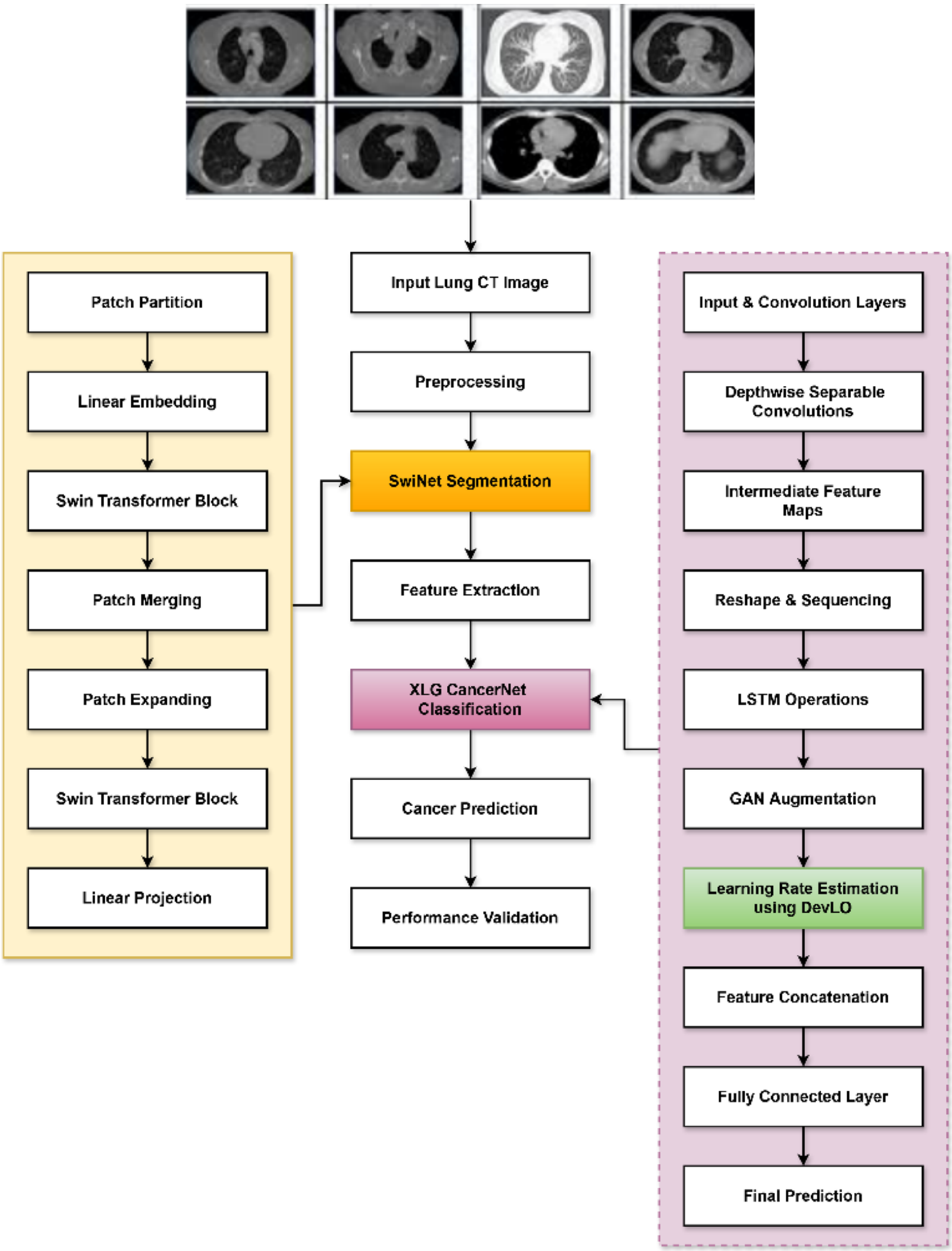


**Fig. 1**. Overall flow of the proposed CanNS model.

feature extraction and classification by XLG CancerNet, and optimally efficient learning rate provided by DevLO finally make this a very strong tool in cancer detection. This framework enhances not only the accuracy of the diagnosis but also helps in developing smarter and more powerful systems for the detection of cancer, which has considerable advantages over the existing techniques and opens up scope for further advancements in the field of medical imaging and disease diagnosis methodology.

Outstanding in capturing hierarchical and multi-scale features in medical imaging, the adoption of Swin-Transformer UNet-or SwiNet-is done in the CanNS framework. For the detection of lung cancer, a critical issue that is in dire need, SwiNet is the efficient integration model enjoying the power of Swin-Transformer with a shifted window-based self-attention mechanism. The proposed mechanism may also give full allowance to capture both the local and global contextual information of CT scan images more efficiently, quite vital when dealing with complex and heterogeneous characteristics in medical images. First proposed for segmentation, the UNet architecture gives complementary insights into the Swin-Transformer attention-based feature extraction due to its symmetric encoder-decoder structure for a pixel-level accurate segmentation of the lung, nodules, and all relevant structures. By focusing Swin-Transformer on diverse scales of information and merging the idea of UNet, which incorporates an encoder-decoder model architecture designed for efficient image segmentation, SwiNet works appropriately with respect to cancer diagnosis of lungs because it highly effectively identifies and segments a medical image containing a region of interest. Besides, Swin-Transformer flexibility can scale up very effectively to large and diverse datasets; hence, it is chosen for this study where the accuracy and robustness across a wide range of patient demographics and clinical conditions are so crucial.

One of the other important decisions involved in this framework was the use of XLG CancerNet, Xception-LSTM GAN for CancerNet. This architecture and reasoning are based on basic rationales that have supported the need for advanced feature extraction, temporal modeling, and data augmentation techniques which raise the accuracy of classification. For feature extraction from the segmented images provided by SwiNet, the Xception model will be used since it is a CNN efficient in the capture of high-order features. This model works, in particular, very well on medical imaging, since it was able to deal with such complexity of features provided by so many different textures, densities, and patterns given, for example, by CT. Context becomes of particular use while analyzing images taken from different angles or across multiple slices of a CT scan. Moreover, GAN for data augmentation is of great importance in medical imaging because getting large annotated datasets is one of the big challenges.

In deep learning, model performance is very sensitive to which kind of optimizer is chosen or its parameters such as the learning rate. It is this adaptive search that will ensure the model evades common pitfalls, such as local minima and overfitting, that severely hamper the performance of deep learning models, especially when their applications involve tasks as complex as medical image analysis. Stability could be one of the possible reasons why DevLO has utilized TDO, while the fact that DevLO belongs to an explorative algorithm in general helps this algorithm converge more efficiently toward optimum by balancing exploration and exploitation, which is more important and critical in the context of this contribution of the CanNS framework aimed at the optimization of the SwiNet and XSL-G cancerNetwork which had to be generalizable by obtaining high performance on various conditions and datasets. This will, in turn, enable DevLO to allow the model to be trained with more accuracy and computational efficiency, which is very crucial in practical healthcare scenarios. These segments together work to address unique challenges associated with medical image analysis and provide a very effective framework for the detection of lung cancer in an effective, efficient, and scalable manner.

Besides clinical conditions, scalability and adaptability on larger and more diverse datasets are very important to consider in order to see applicability in real-world settings. Advanced deep learning models and optimization techniques to handle high-dimensional data with flexibility in different clinical settings have been some of the major strengths in the design of the CanNS Framework. It allows the Swin-Transformer UNet segmentation model to handle huge volumes of CT scan images using an attention mechanism and hierarchical features extracted from the regions of interest. This ultimately makes the network adapt to several changes in image quality, variations in patient anatomy, and numerous pathologies that may appear in the lung. Apart from that, the generalization capability for various clinical datasets has been further advanced in the proposed XLG CancerNet by including the powerful Xception for feature extraction, LSTM network modeling of temporal patterns, and GAN network for augmentation. GAN-based augmentation enables large numbers of synthetic samples for training the models, hence enhancing model robustness for tackling issues with an imbalanced class distribution or some kind of rare diseases. In this view, the DevLO methodology plays a very important role within the framework for the adaptation of models to different clinical conditions via dynamic optimization of the learning rate and updating the parameters in such a way that allows an adaptive search procedure.

This becomes very important if later the framework needs to be applied on newer and bigger datasets with different characteristics since this retains the accuracy and robustness of the model while keeping the computational overhead much lower. This fine-tuning of model parameters in DevLO for the conditions involves an easy-adaptation framework on a number of scanning conditions, various scanner types, and even different patient demographics that will be diversified further within the clinical setup. High scalability can also be done using computational efficiency to perform large-scale datasets with large timescales without compromising any performances. Continual learning and fine-tuning would be very important in those cases where either the size of the dataset increases or new clinical conditions are met with the CanNS Framework. This means that the architecture of this framework will let the model integrate other data modalities, like various imaging techniques or even multi-center data, which further lets the model adapt to different clinical conditions. By using transfer learning and periodic retraining of the model with new data, the model is able to continuously evolve towards the demands of changing clinical practice. In all, the CanNS framework is perfectly positioned for scaling up larger and more diverse datasets, adaptable to a wide variety of clinical conditions, hence promising for real-world healthcare use in lung cancer detection.

### A. *Data source and data preprocessing*

The dataset to be used for this project, a publicly available one on Kaggle, the Lung CT Image Dataset, has been widely recognized for its relevance in lung cancer research. Lung cancer is one of the most important causes of deaths that have something to do with cancer, and diagnosis at an early and appropriate stage is crucial in giving chances to prolong the life of an affected patient. The dataset consists of high-resolution lung CT images, highly desirable when developing robust diagnostic models, in particular with the usage of cutting-edge deep learning and transformer-based architectures. This will be a correct dataset due to its variability and quality, hence reliable in training and testing the Cancer Nexus Synergy (CanNS) framework in an integrative manner. All CT images were resized to a standard dimension initially so that all images in the dataset would be uniform and, therefore, compatible with the Swin-Transformer UNet (SwiNet) and XLG CancerNet models. These images were normalized to remove the variation in intensity due to different scanning conditions, enhancing robustness by removing bias in models. As mentioned earlier, the following are the different augmentations made on this data in terms of random rotations, random flips, random scaling, and adding random noise to enhance artificially the dataset size that helped in overcoming some overfitting issues while improving model generalization capabilities. These segmentations were preprocessed by the SwiNet model to outline the lung regions from CT scans and remove unwanted background information. Pre-processed segmented output is used for classification by XLG CancerNet to identify correct cancer cases.

The dataset used in this work contains 1,190 CT scan images that are slices from 110 cases, clinically diagnosed into three classes: normal, benign, and malignant. Of these, 55 cases were classified as normal and contributed 600 images, 15 cases were diagnosed as benign and contributed 190 images, and 40 cases were identified as malignant and contributed 400 images. Each scan consists of 80 to 200 slices, all representing the human chest from different sides and angles. All images were de-identified before analysis, and the study was approved by the institutional review board of participating medical centers, with written consent waived by the oversight review board. The dataset details are provided in Table 1.

### B. *Swin-Transformer UNet (SwiNet) model for segmentation*

In the context of medical image analysis which includes, for instance, cancer detection, preprocessing is a fundamental task. The purpose is to optimize the images and normalize them so as to feed the segmentation model with proper images. It often includes normalization in which the pixel intensities are standardized with reference to a common range thus reducing the impacts of difference in lighting. Another form of property to the data is through the use of augmentation techniques that include rotation, flipping, and zooming since it will improve the understanding of the model to other variations that it can encounter. The other a crucial step is denoising especially in medical images since noise and artifacts may hamper vision of features. It is common to use techniques such as Gaussian filtering, or median filtering to eliminate this kind of noise which will better bring out structures that the model needs to learn from. This way we guarantee that the images that are fed into the segmentation model are of the best quality, standard and hold the best capabilities of identifying cancerous regions and segmenting them. In order to address the limitations of the prior models, the unique and novel SwiNet model is used in this work for solving medical image segmentation problem, aimed at cancer detection.

The model begins with segmenting the input image into overlapping patches where none of those patches overlap with each other. This first step is essential because it makes it easier for the model to analyze sections of an image – and, consequently, because it enables finer levels of analysis on these particular regions. The developments of vision transformer models like Swin Transformer give many benefits in the image segmentation task, especially in medical imaging. One availed benefit of the three is that they embrace global contextual information. The CNNs have issues with their receptive fields and indeed, with modeling long distance dependencies and relations throughout the image. Vision Transformers, on the other hand use multi-head attention mechanisms which enable each of the patches to look at all other patches that are present in the image. Another good aspect of vision transformers is that it allows for hierarchical feature extraction. The image in models like the Swin Transformer is processed at different resolutions at once ensuring that the necessary fine and the course features are captured. This form of representation by multi-scale is highly useful in medical image segmentation as lesion or tumor size might be very large or very small. Vision Transformers are also highly flexible and can input images of virtually any size and resolution across the commonly used medical imaging modalities varying from high-resolution histopathology whole-slide images to lower resolution MRI scans. Moreover, Vision Transformers have lesser inductive bias in comparison to CNNs. It could be very useful in medical imaging when the features of interest, such as tumors, cannot be placed or may be located in different form or shapes.

The proposed SwiNet model is a combination of Vision Transformer and UNet model, thus, has several advantages over the existing models. Among others, one should mention the hierarchical representation of the structure. SwiNet performs well as it works on multiple scales due to the structure of its network and thus considers the high details of the object and its surroundings. This is very crucial especially in medical imaging where localization achieved by outlining the boundaries of an element for instance a tumor and the global structures such as the organ that hosts the element in question are equally vital for segmentation. The SwiNet also utilizes the local-global attention which is able to combine the feature extraction of CNNs for detailed tasks with the Vision Transformers' capacity for global context modeling. This is an advantage for SwiNet since the site can perform well when both local and global information is crucial to finish a task such as discernment of tumor boundaries. Moreover, the patch merging and expanding operations in SwiNet also solves the issues of computational complexity and memory requirement for the model so that the model can process high-resolution medical imagery without significant computational cost. SwiNet also has some other specialized benefits, which stem from the fact that it employs skip connections, in the same manner as UNet does. Such skip connections assist in maintaining the spatial resolution and high details that are very essential in medical image segmentation tasks. As a result, SwiNet is established to have competitiveness and superiority to traditional models that only employs the CNN or Transformers for cancer image segmentation.

| Classes | No of cases | No of images |
|---|---|---|
| Normal | 55 | 600 |
| Benign | 15 | 190 |
| Malignant | 40 | 400 |
| Total | 110 | 1190 |

**Table 1**. Dataset details.

Therefore, the proposed SwiNet model's approach of integrating Swin Transformer blocks with UNet for medical image segmentation is unique. This helps in getting the benefits of both types of architectures and hence train a model which can predict with accuracy the detailed local features and at the same time contains long-range dependencies and global context. The proposed methodology can be useful in the further development of the method of medical image segmentation. Figure 2 mentions the architecture model of proposed SwiNet model.

It presents a model of applying the Vision Transformers with the state-of-art performance and the UNet structure that has been recently demonstrated to be efficient in the tasks of cancer lesions detection and segmentation in medical images.

In this technique, the segmentation process is initialized by splitting the input image having size of $\langle \times \sqsupseteq \times \rfloor$, where $\langle$ indicates the height, $\sqsupseteq$ denotes the weight and $\rfloor$ represents the number of channels. At first, the image is partitioned into non-overlapping patches, which is the main operation of segmentation. During this process, the image is transformed into smaller and manageable portions that enables the model to completely focus on the localized features while preserving it spatial structures. This operation is mathematically represented in below:

$$m_p = Reshape(\mathbb{I},\ (\aleph\ ,\ \mathfrak{p} \times\ \mathfrak{p} \times\ \rfloor)) \tag{1}$$

Where, $\mathbb{I}$ indicates the input image, $\aleph$ represents the total number of patches, $\mathfrak{p} \times\ \mathfrak{p}$ denotes the dimension of each patch and $m_p$ is the image patch. This partitioning helps fine-grained feature extraction, which means that the model will capture small changes in the image, something very relevant in medical imaging where slight abnormalities may carcinify a disease. Consequently, the encoder will transform these input patches through a number of successive stages of processing into a lower-dimensional space while progressively extracting higher-level features. During linear embedding, each patch $m_p$ is transformed into a lower dimensional feature space as represented in the following model:

$$\delta\ = m_p \varpi_{\ e} + \lfloor_e \tag{2}$$

Where, $\varpi_{\ e}$ indicates the weight value and $\lfloor_e$ is the bias value. These patched embedding are then fed into Swin Transformer blocks to extract local and global contextual information through self-attention mechanisms and feed-forward networks. It is a mechanism enabling every patch to attend to all the other patches, hence recovering long-range dependencies and contextual relationships. Here, self-attention allows every patch to attend to all others in modeling long-range dependencies and contextual relationships. It is mathematically represented in the following model:

$$Attention\ (\mathfrak{X}, \mathbb{K}, \mathfrak{E}) = Softmax \frac{\mathfrak{X}\mathbb{K}^T}{\sqrt{\mathfrak{d}_{\mathbb{K}}}} \mathfrak{E} \tag{3}$$

Where, $\mathfrak{X}$ indicates query, $\mathbb{K}$ represents the key value, $\mathfrak{E}$ is the value, and $\mathfrak{d}_{\mathbb{K}}$ denotes the dimensionality of key vectors. As a consequence of this, the layer normalization and residual connection operations are performed to stabilize the training and validation operations as represented in below:

$$\dashv_1 = LinNorm\,(\dashv_0) + MHSA\,(LinNorm(\,\dashv_0\,)) \tag{4}$$

$$\dashv_2 = LinNorm\,(\dashv_1) + MHSA\,(LinNorm(\,\dashv_1\,)) \tag{5}$$

Where, $\dashv_1$ indicates the output from first MHSA, and $\dashv_2$ indicates the intermediate feature map after FFN. Then, the patch merging operation is performed to minimize the total number of patches by adding it with adjacent patches, which significantly supports to down sample the features while preserving the most useful information. The following model can be used to perform patching merging operation:

$$\dashv_x = Concat\,(\dashv_2,\ \dashv_{2+1})\,\varpi_{\ \dashv} \tag{6}$$

Where, the patches $\dashv_x$ are concatenated and linear transformed for minimizing the dimensionality. Then, the encoder can use multiple stages of Swin transformer blocks that are integrated with patch merging for getting hierarchical representations of image at multiple scales. It supports to improve the spatial resolution of feature maps, and helps to effectively upsample the image back to the original input image. The patch expanding operation is mathematically represented in below:

$$\dashv_\rceil = Expand\,(\dashv_{Enc}) \tag{7}$$

Where, $\dashv_\rceil$ is the expanded feature map, $\dashv_{Enc}$ indicates the encoded feature map. Similar to encoding operation, the decoder block also uses the swin transformer blocks in order to refine the upsampled features, which preserves both local and global contexts. This operation is mathematically represented in below:

$$\dashv_{\mathfrak{d}1} = LinNorm\left(\dashv_\rceil\right) + MHSA\,(LinNorm(\,\dashv_\rceil\,)) \tag{8}$$

$$\dashv_{\mathfrak{d}2} = LinNorm\,(\dashv_{\mathfrak{d}1}) + MHSA\,(LinNorm(\,\dashv_{\mathfrak{d}1}\,)) \tag{9}$$

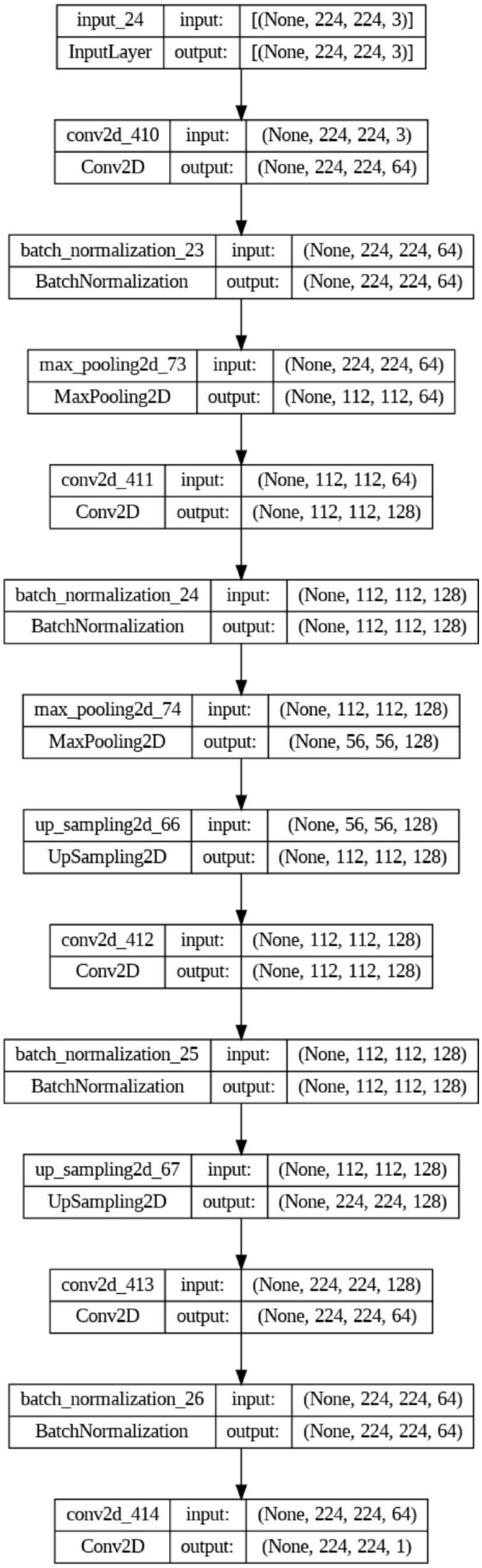


**Fig. 2**. Architecture model of proposed SwiNet model.

Where, $\dashv_{\partial 1}$ denotes the intermediate feature map in the decoding side, $MHSA(.)$ indicates the multi-head self-attention module, and $LinNorm\,(.)$ Moreover, the skip connections are used to integrate the feature maps from encoding and decoding stages for maintaining high resolution image details, which effectively maintains the spatial information for determining an effective image segmentation. This operation is mathematically illustrated in below:

$$\dashv_{skip} = Concate(\dashv_{Enc},\ \dashv_{Dec}) \tag{10}$$

Furthermore, the linear projection operation is performed to refine all features for determining the output segmentation map as defined in the following equation:

$$\mathcal{S} = \dashv_{Dec}\varpi_{\ p} + \lfloor_p \tag{11}$$

Where, $\mathcal{S}$ represents the final segmented output, $\varpi_{\ p}$ and $\lfloor_p$ indicates the weight and bias values of the projected layer. By using the proposed SwiNet model, the medical image segmentation is effectively performed in the proposed work, which supports to gain an improved cancer detection results.

### C. *Xception-LSTM GAN (XLG) CancerNet based classification*

After separating the important areas from the medical images using SwinNet model, the deep classification of these areas is achieved by the XLG CancerNet. It is a hybrid architecture that integrates three powerful deep learning techniques: Xception Net is one of the efficient residual based model developed by the hybridization of CNN, Deep LSTM and Generative Adversarial Networks (GAN). All these components have critical roles of improving the model's efficiency in identifying cancerous tissues and classifying them in medical images. The segmented patches of the medical images are, to begin with, passed to the Xception Net, which is an enhanced CNN architecture that is widely renowned for unveiling better features from the input data. Moreover, the Xception Net employs such a network architecture feature as depthwise separable convolutions which replace standard convolutions and consist of depthwise convolutions and point wise ones. It lets the network learn precise spatial relationships and elaborate texture patterns associated with cancer niches in medical images. The mathematical formulation entails the use of several convolutions and non-linearity activists which transforms the input image to several feature maps. After the features are extracted through the Xception Net they are re-sized and followed by a Deep LSTM network. LSTMs are members of a subset of neural networks called RNN that are developed to learn sequential data and temporal dependencies. When working with nine medical images, LSTM can learn the order of the features extracted from the images and therefore, the evolution and dependency of the patterns that may reveal the development of cancer. The hidden states of the LSTM elements are updated thusly and these are capable of preserving long term dependencies of the feature sequence. To improve the model's ability to withstand certain noise level and data variance, the GAN is integrated into the ML algorithm. GANs are networks, which include a generator and a discriminator network essentially operating in the same environment. The generator synthesizes fake medical images employed to extend the set of labeled images, since the availability of the labeled data is a problem. This augmentation widens the network's learning scope to various features other than the specific image reducing the chances of over fitting and making it easier for the model to predict and recognize similar images in the unseen data set.

From the GAN, the Xception Net features passed through LSTM, and the classification output of the GAN made up a perfect feature set for classification. The separated features are provided to fully connected layers where every layer performs a linear transformation then applies non-linearity to improve the features. The last layer uses the softmax activation function to provide the probabilities of each class with the presence of cancerous tissues. Regarding, the novelty of the XLG CancerNet, it is necessary to mention that this tool applies the three most effective methods all combined during the basis of the model. By separating it convolution wise depending on the depth of Xception Net, it is possible to present a precise and efficient feature extraction. Deep LSTM is useful in capturing temporal information, as well as in creating a sequence of patterns since the formation of cancer changes in tissues is a progressive process. Subsequently, GANs are further used to enhance the experience of the training method whereby the algorithm continuously creates the realistic data of high quality that help enhance the stability of the model and overall generalizing capability. This makes the proposed hybrid model suitable for cancer identification in medical images because; first it has strong identification and detailed analysis on the input data mainly on spatial and temporal aspects which are vital for diagnostic outcomes. The Fig. 3. Represents the architecture of proposed XLG-CancerNet model.

Secondly, the kind of dispersion influences the training occupations, the application of GANs also reduces the problem of the scarcity of annotated medical data to improve the model's performance by expanding the training data set. Finally, the use of these unique methods in one model increases the model's effectiveness and adaptability to the degree of irregularity and diversity inherent in medical imagery information. Overall, the developed XLG CancerNet is the next generational hybrid model, which is the combination of Xception Net, Deep LSTM, and GAN that are powerful in detecting cancer images from the provided medical image dataset. Since, the proposed XLG CancerNet utilizes an efficient feature extraction, sequential pattern modeling, and the possibility of data augmentation, as significant for precise cancer diagnosis and therefore improved clinical prognosis.

At first, the SwinNet model segments the input medical images before passing the segmented output into the XLG CancerNet. The segmented areas likely include cancerous tissues that would then be fed forward into the XLG CancerNet for classification. During this process, the depthwise separable convolution operations are performed for an effective feature extraction, which is mathematically represented in below:

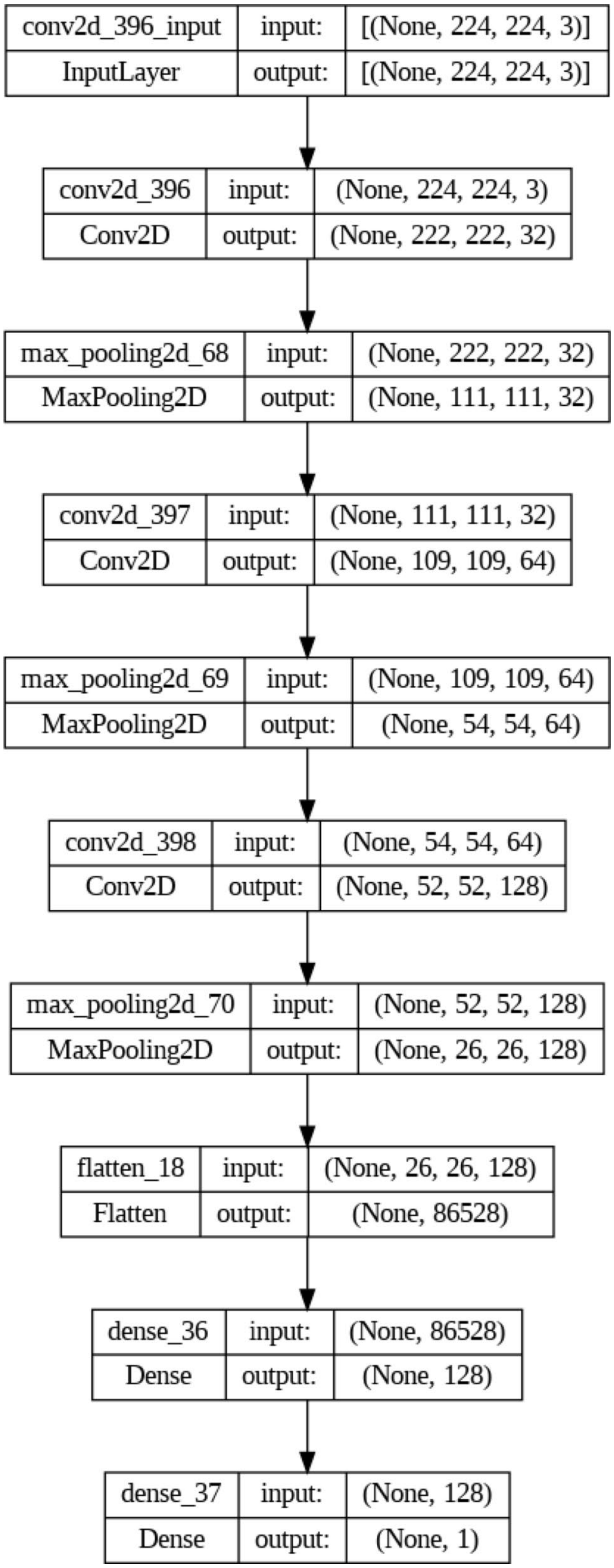


**Fig. 3**. Architecture of proposed XLG-CancerNet model.

$$\sim_1 = Bat_{Norm}(Conv2D\,(\sim_0, \mathfrak{w}_1) + \mathfrak{P}_1) \tag{12}$$

Where, $\sim_0$ indicates the input image passed to the classifier, $\mathfrak{w}_1$ and $\mathfrak{P}_1$ are the weight and bias values.

$$\sim_{dep} = Bat_{Norm}\,(DepthwiseConv2d(\,(\sim_1, \mathfrak{w}_{dep}) + \mathfrak{P}_{dep} \tag{13}$$

$$\sim_{poi} = Bat_{Norm}\,(DepthwiseConv2d(\,(\sim_{dep}, \mathfrak{w}_{poi}) + \mathfrak{P}_{poi} \tag{14}$$

Where, $\sim_{dep}$ indicates the output of depthwise convolution and $\sim_{poi}$ represents the output of pointwise convolution. After performing depthwise convolutions and non-linear activation processes, the intermediate feature maps are obtained as represented in below:

$$\sim_{\{} = Xception\,(\sim_0) \tag{15}$$

Where, $\sim_{\{}$ indicates the feature maps. As a consequence of this, the temporal pattern modeling is performed with the use of deep LSTM, where the obtained features are reshaped for estimating the temporal dependencies as mathematically illustrated in below:

$$\sim_{\mathfrak{r}} = Reshape\,(\sim_{\{},\ \mathfrak{L},\ \mathcal{D}) \tag{16}$$

Where, $\mathfrak{L}$ indicates the sequence length and $\mathcal{D}$ represents the feature dimensionality. Moreover, the LSTM operations including input gate, forget gate, output gate, cell state, and hidden state estimation are performed, and the final LSTM output is obtained as below:

$$\mathfrak{h}_{\{} = LSTM_{out}\,(\sim_{\mathfrak{r}}) \tag{17}$$

Where, $\mathfrak{h}_{\{}$ indicates the output hidden state. Then, the augmentation is performed for getting the realistic images as illustrated in the following models:

$$\varsigma_{\,v} = \rho_{\,g}\,(g) \tag{18}$$

$$\sim_v = Gen_N(\varsigma_{\,v}, \mathfrak{w}_v, \mathfrak{P}_v) \tag{19}$$

Where, $\varsigma_{\,v}$ indicates the random noise vector obtained from prior distribution $\rho_{\,g}\,(g)$, and $\mathfrak{w}_v, \mathfrak{P}_v$ are the weight and bias values respectively. Then, the generated images are obtained as illustrated in the following equations:

$$Dis_N\,(\sim_{\mathfrak{r}},\ \mathfrak{w}_{dep}, \mathfrak{P}_{dep}) = \phi\,(\lambda_{\,dis} \times\ \mathfrak{w}_{dep} \times\ \sim_{\mathfrak{r}} + \mathfrak{P}_{dep}) \tag{20}$$

$$Dis_N\,(\sim_v,\ \mathfrak{w}_{dep}, \mathfrak{P}_{dep}) = \phi\,(\lambda_{\,dis} \times\ \mathfrak{w}_{dep} \times\ \sim_v + \mathfrak{P}_{dep}) \tag{21}$$

Where, $\sim_{\mathfrak{r}}$ is the obtained image, $\sim_v$ represents the generated image, and $\lambda_{\,dis}$ indicates the learning rate that is estimated with the use DevLO optimization algorithm. Then, the adversarial loss function is estimated based on the following equation:

$$\mathbb{L}_{Dis} = -\mathcal{E}_{\sim_{\mathfrak{r}} \sim \rho_{\,data}}\,[\log Dis\,(\sim_{\mathfrak{r}})] - \mathcal{E}_{\sim_v \sim \rho_{\,g}(g)}\,[\log(1 - Dis(Gen\,(\sim_{\mathfrak{r}})))] \tag{22}$$

$$\mathbb{L}_{Gen} = -\mathcal{E}_{\sim_v \sim \rho_{\,g}(g)}\,[\log Dis\,(Gen\,(\sim_{\mathfrak{r}}))] \tag{23}$$

Where, $Dis(Gen\,(\sim_{\mathfrak{r}})$ indicates the discriminator probability output, and $Dis\,(\sim_{\mathfrak{r}})$ is the discriminator output. Then, the features of Xception Net, deep LSTM and GAN models are concatenated together to obtain the final feature vector as shown in below:

$$\sim_c = Concate(\sim_{\{}, \mathfrak{h}_{\{}) \tag{24}$$

Then, the combined features are all passed to the fully connected layer for getting the final predicted result as defined in the following equation:

$$f_{out} = Softmax(\mathfrak{w}_n \times\ \sim_{n-1} + \mathfrak{P}_n) \tag{25}$$

Where, $f_{out}$ indicates the final output.

```
Input:    Input medical image $\mathfrak{I}$;
Output:   Predicted results with probability;
Step 1:   Initialize the parameters of XceptionNet, deep LSTM and GAN;
Step 2:   For each epoch in N
              For each batch in data loader
Step 3:               Perform segmentation using SwinNet model;
Step 4:               Obtain the segmented image $\mathbb{s}_0$ from input $\mathfrak{I}$;
Step 5:           Extract features $\mathbb{s}_\mathfrak{f}$ from the image $\mathbb{s}_0$ using Xception Net as mentioned in equ (14);
Step 6:           Perform temporal pattern modeling using LSTM;
Step 7:           Reshape features $\mathbb{s}_\mathfrak{f}$ into the sequences of $\mathbb{s}_\mathfrak{r}$ based on equ (15);
Step 8:           Find temporal dependencies using LSTM for getting $\mathfrak{h}_\mathfrak{f}$ based on equ (16);
Step 9:           Perform data augmentation using GAN, in which the noise $\varsigma_v$ is generated from
                  prior distribution $\rho_g$ using equ (17);
Step 10:          Generate an augmented image $\mathbb{s}_v$ using $\varsigma_v$ based on equ (18);
Step 11:          Estimate the discriminator outputs $Dis_N$ using equ (19) and (20);
Step 12:          Calculate the adversarial loss functions $\mathbb{L}_{Dis}$ and $\mathbb{L}_{Gen}$ using equ (21) and (22);
Step 13:          Concatenate all features $\mathbb{s}_\mathfrak{f}$ with $\mathfrak{h}_\mathfrak{f}$ to get $\mathbb{s}_c$ as mentioned in equ (23);
Step 14:              The final classified output is predicted using the fully connected layer, and the
                      softmax function is estimated to get $f_{out}$ using equ (24);
              End for;
Step 15:  End for;
```

**Algorithm 1**. XLG CancerNet model.

### D. *Devilish Levy Optimization (DevLO) for parameter tuning*

The Devilish Levy Optimization (DevLO) technique is mainly in this study for optimally estimating the learning rate of the XLG-CancerNet model for cancer diagnosis using medical images. The learning rate occupies an important place in the deep learning since the mentioned gradient is connected with the step during the use of gradient descent, and, therefore, with the speed of convergence of the model to the optimal solution. To a certain extent, another significant issue of the present work is the choice of Adam's optimizer parameters during training; in particular, when using models that include multiple sub models: The Xception structure is used for feature extraction, LSTM for temporal pattern modeling, and GAN for data augmentation for the further application of the XLG CancerNet. Therefore, the enhancement of the learning rate still holds one of the key practices in the context of XLG CancerNet. Selecting a proper learning rate helps the model to get well trained during the training session and at the same time discourages the model to select very high convergence rates. On the other hand, if the learning rate is set to large the model can move over the best optimum to the worst and thus will continue the process in circles, meaning low performance. On a low note, learning rate that is set too low will take long times to converge thus prolonging the time to be trained. Consequently, the challenge is aimed at the kind of way that DevLO employs the adaptive search procedure which hinges on TDO together with the explorative property of Levy flight. As can be seen, the major components of DevLO are TDO and Levy flight which were never incorporated together before. This real-time experience in learning, as seen in TDO which is based on hunting techniques of Tasmanian devils, therefore provides a foundation for switching the learning rates depending on the feedback on the performance. It has exploration of different areas at the same rate with the exploitation of rich areas at the same rate so that the learning rate values are conditioned to be improved in the course of training. While Levy flight returns stochasticity in random walks which perform large jumps, it guides DevLO to search a large space of learning rate efficiently.

Deep learning model performance and convergence heavily depend on some hyperparameter choices in handling very complex tasks; for example, detecting lung cancer based on medical images. Of these, the learning rate is usually the most important hyperparameter that controls how much a model would change the weights in one training step. A high learning rate leads to overshooting over the optimum, while a low one results in very slow convergence and can make convergence much slower. The batch size can be defined as the number of samples in one batch, processed before updating model weights. Indeed, DevLO seems pretty appropriate for optimizing those hyperparameters in order to guarantee that the optimal models of both segmentation, SwiNet, and classification, XLG CancerNet, make good performances without being led by some common pitfalls such as overfitting and/or premature convergence.

DevLO implements an adaptive search procedure that shall smartly adapt the hyperparameters in view of better model performance. The most exciting point in DevLO deals with the adoption of the Levy-flight mechanism for the effective exploration of solution space. This can prevent gradient-based search methods from getting stuck into local minima, as Lévy flight enables DevLO to travel large expanses of the search space toward the most promising areas for improvements, whereas optimization of model hyperparameters has to be performed. To this end, DevLO is particularly fitted to solve such complex, high-dimensional search spaces that deep learning models entail with several hyperparameters having to be tuned simultaneously. The DevLO algorithm, while performing fine-tuning in learning rates, will definitely tune each layer of this multi-layer XLG CancerNet model in a differential manner during training. DevLO makes sure that dynamic adjustment of the learning rate within diverse layers and sub-models results in advanced convergence of the model to the optimal solution in context to the detection of lung cancer.

Besides, DevLO is not restricted to only optimizing the learning rate but extends capability to other hyperparameters participating in the training, that include momentum, decay rates, and activation functions of network layers. Similarly, there are also decay rates: controlling the gradual decrease of learning rates in training, also featured much in the process of fine-turning model performances to get closer to a perfect solution. DevLO aims at the best decaying rate that can balance fast convergence and not being caught into poor local minima.

It is also important to underline that DevLO has some nice features in comparison with other conventional optimization techniques or other general search methods such as grid search. It can regulate the changes in the training condition of XLG CancerNet, and it can deal with the issue linked to learning rate optimization conveniently and effectively. Therefore, the addition of TDO and Levy flight into DevLO enhances the discovery of proper updated learning rate to further speed, accurate convergence, and efficiency of the model for all the training datasets. Because of this adaptability and exploration nature, DevLO is an effective approach for the additional improvement of deep learning models like XLG CancerNet in the terms of accuracy and efficiency for the domains, such as medical image analysis for cancer detection. In this technique, the obtained feature set, number of epochs and batch size with initial learning rate are taken as the input parameters, and the estimated optimal value is returned as the output.

The population initialization process is carried out at first as shown in the following equation:

$$M_{ij} = m_j^{mn} + \xi \times (m_j^{mx} - m_j^{mn}) \tag{26}$$

Where, $i = 1,2,3\ldots X$ and $j = 1,2,3\ldots y$, $X$ indicates the number of Tasmanian devils, $y$ is the total number of variables, $\xi$ denotes the random number, $m_j^{mn}$ and $m_j^{mx}$ are the lower and upper limits. In this strategy, the Tasmanian devils searching for their natural food, carrion, scattered over some environment. Here, each Tasmanian devil represents an agent that optimizes a certain problem and explores different parts of a search space in search of an optimal solution. "Carrion" positions simply define the possible fields in a search space or, more interestingly, point out some promising areas. In the process of optimization, each devil randomly selects one of these locations where it has the carrion to further investigate in an attempt to improve its current position in the search space. This random selection makes a great diversity in exploration that can lead to the discovery of new solutions probably much better than those already found on this landscape of the search. It is mathematically represented in below:

$$\mathtt{C}_i = M_h,\ i = 1,\ 2,\ \ldots X,\ h \in \{1,2\ldots X | h \neq i\} \tag{27}$$

Where, $\mathtt{C}_i$ indicates the selected carrion, and $h$ represents the randomly chosen number. As a consequence of this, the position updation is performed based on the following equation:

$$m_{i,j}^{new,\,P1} = \begin{cases} m_{ij} + \partial \times (\mathfrak{c}_{ij} - \mathfrak{T} \times m_{ij}) & \{\mathtt{C}_i < \{_i \\ m_{ij} + \partial \times (m_{ij} - \mathfrak{c}_{ij}) & Otherwise \end{cases} \tag{28}$$

$$M_{ij} = \begin{cases} M_i^{new,P1}, & \{_i^{new,P1} < \{_i \\ M_i, & Otherwise \end{cases} \tag{29}$$

Where, $M_i^{new,P1}$ indicates the new position of Tasmanian devil, $\{_i^{new,P1}$ represents the objective function new value, $\{_{\mathtt{C}_i}$ defines the objective function for the selected carrion, $\partial$ and $\mathfrak{T}$ are the random numbers. During the exploitation stage, the member *h* is chosen at random according to the position of prey as defined in the following model:

$$\mathcal{Q}_i = M_h,\ i = 1,\ 2,\ \ldots\ X,\ h \in \{1{,}2 \ldots\ X|h \neq\ i\} \tag{30}$$

Then, the newly updated position of Tasmanian devil is estimated as shown in below:

$$m_{i,j}^{new,\ P2} = \begin{cases} m_{ij} + \partial\ \times\ (q_{ij} - \mathfrak{T} \times\ m_{ij}) & \{_{\mathcal{Q}_i} < \{_i \\ m_{ij} + \partial\ \times\ (m_{ij} - q_{ij}) & Otherwise \end{cases} \tag{31}$$

$$M_i = \begin{cases} M_i^{new,P2}, & \{_i^{new,P2} < \{_i \\ M_i, & Otherwise \end{cases} \tag{32}$$

Where, $M_i^{new,P2}$ indicates the new position, and $\{_i^{new,P2}$ represents the objective function of the selected prey. According to the chasing process, the new location is updated with the neighborhood as shown in the following equation:

$$\mathcal{R} = 0.01 \times\ (1 - \frac{itr}{Mx_{itr}}) \tag{33}$$

$$m_{i,j}^{new} = m_{ij} + (2\alpha\ - 1) \times\ \mathcal{R} \times\ m_{ij} \tag{34}$$

$$M_i = \begin{cases} M_i^{new}, & \{_i^{new} < \{_i \\ M_i, & Otherwise \end{cases} \tag{35}$$

$$\alpha\ = e^{\left(-\tau\ |u|^{\vartheta}\ \right)},\ \vartheta\ \in\ [0{,}2] \tag{36}$$

Where, $\mathcal{R}$ indicates the neighborhood radius value, $itr$ represents the current iteration, $Mx_{itr}$ is the maximum number of iterations, $\alpha$ indicates the random value computed using Levy distribution strategy, $\vartheta$ is the distribution index and $\tau$ scaling parameter.

The proposed DevLO-or Devilish Levy Optimization-lays in the fact that it gives an efficient estimate of the learning rate of the XLG-CancerNet model adaptively, which is needed for the appropriate diagnosis of cancer from any medical image. The learning rate is a very important hyperparameter of the deep learning model, since it decides the speed of convergence of any model while training. It can result in overshooting the optimal point or, even worse, failing to converge into an optimal solution for better performance. Traditional optimization methods are gradient descent and Adam optimizer. Both of these have limitations for the fine-tuning of hyperparameters of learning rates of complex models with several sub-models, like XLG-CancerNet. Since this is a multi-component framework, choosing an appropriate learning rate is not that easy; hence, it needs an optimization approach which can go well with complications brought forth by the combined models. Standard optimization methods, such as Adam, very often will be at least insufficient on their own, since they rely on pre-defined learning rate schedules that cannot adapt to the intricacies of deep architectures with lots of interacting components. DevLO was developed to overcome these shortcomings. Unlike other traditional optimization techniques, DevLO has an adaptive search procedure with the strengths of two key elements in its combination: TDO-Tuned Differential Operator and the Levy flight's explorative capability. Levy flight enables more careful exploration of the search space than the algorithm can avoid or locate the global optimum at the expense of local optima. In other words, such features of DevLO allow the learning rate during training to be changed dynamically in such a way that the convergence speed, together with model accuracy, is optimized without overshooting the optimal solution. The classic ones include gradient descent and the Adam optimizer, neither of which is as adaptive to complex, multi-dimensional problems. Unfortunately, most of them are dependent on some fixed or pre-defined learning rate which, in consequence, can get the optimum learning rate for some phases of training. DevLO combines the power of Levy flight and TDO for its adaptive driving in better explorations within the parameter space for an optimum learning rate in pursuit of training phases. This will make them suitable for models such as XLG-CancerNet, which depend upon these kinds of fine-grain optimization among multi-components. This leads to faster convergence, higher accuracy, and finally better performance in detecting cancers.

## Results and discussion

This section validates the performance outcomes of the proposed framework of Cancer Nexus Synergy and further validates the comparative results using different evaluation metrics. The model is thus developed for efficient and strong cancer diagnosis and hence rigorously tested and evaluated, aiming to prove its effectiveness. Various metrics have been used to measure performance, such as accuracy, sensitivity, specificity, and AUC. Comparisons have also been made with respect to some other existing models, elaborating on the improvements and advantages that the introduction of CanNS brought. Each evaluation metric contributes to an all-rounded understanding of the strengths of the model and possible perspectives of areas for improvement. Such results are the outcome of intensive experimentation in trying to test whether the proposed framework—CanNS—would finally improve the accuracy and reliability of cancer diagnosis. It also brings out the power of the framework in solving intricacies and challenges taken to represent problems in medical image analysis. In this regard, the proposed model is tested with the use of the publicly available Kaggle lung CT image dataset[39]. This can be treated as a comprehensive dataset of lung CT images that would be highly required to test the performance of the CanNS model in a real scenario. The accuracy, sensitivity, and specificity of the model in detecting and diagnosing cancer are checked in this dataset. The great variety of images in the dataset will help view most of the cases and conditions on the model's capability, hence proving to be a robust and reliable model. It involves detailed analysis and comparison with existing models during its validation process, putting forward the improvements and advantages offered by the CanNS framework. This CanNS model applied to strict validation possesses huge potential to rise to very high standards in cancer diagnoses, producing reliable and accurate results that help medical professionals in making effective decisions.

The Kaggle dataset[40] of lung CT images is an open repository of a large collection of lung CT images used majorly for training and validation of even deeper models of machine learning and deep learning in medical image analysis applications. This dataset contains a strongly diversified computed tomography image dataset of patients at different stages and types of lung conditions, building strong models that generalize well across different patient profiles and manifestations of diseases. Most of the images come distinguished with detailed annotation or labels regarding the presence and location of the nodules or cancerous regions, critical in training supervised learning models and evaluating their performance. High-resolution images ensure that models learn from and operate on fine-grained details essential for the accurate detection and diagnosis of lung cancer. This typically contains types of relevant metadata attached to the images, such as age, sex, and clinical history, providing context data that helps improve model performance. It has a wide span of cases, including healthy lungs and all forms and stages of lung cancer, thus modeling the differences between normal and pathologic conditions. Correspondingly, because it lives in Kaggle, it becomes part of this large community of data scientists and researchers who can work together, share ideas, and use a plethora of pre-trained models and kernels. This dataset shall then be trained with validation and test sets to the standard benchmark for that dataset, providing their evaluation metrics in a comparable format. The resulted models will then be applied in a real clinical setting to aid in the detection, segmentation, and diagnosis of cancer, and further help fine-tune the proposed CanNS model to high standards of accuracy and reliability clinically.

Figure 4 shows the sequential steps of the proposed CanNS framework's flow from input to final classification. On the left is the raw input CT lung scan, which captured the whole of the lung tissue along with all the complexities, variations that usually occur in any medical imaging. This raw image is very important since it provides a baseline for all the processing stages ahead. The second image represents the result after preprocessing. Preprocessing techniques usually work on the reduction of noise, normalization, and enhancement of images to improve their quality before appropriate segmentation and extraction of features. These steps are very important in highlighting relevant structures within the lung, making sure that latter model inputs will be clean and standardized. The third image describes the segmentation done using the SwiNet model. SwiNet is powered by advanced transformer-based architecture that enables efficient identification and isolation of regions of interest, with a great focus on the potential cancerous areas in the lung CT scan. This model is very fine at handling the intricacies of medical images, giving precise and inclusive segmentation. The result of the segmented image shows parts like nodules or tumors in different colors from the rest of the lung tissue, hereby visualizing them and analyzing disjointedly. The last predicted image after the classification using the XLG CancerNet model. This is a deep learning model in whose architecture the use of Xception and LSTM in feature extraction and pattern identification characterizes the presence of cancer. These are input into XLG CancerNet for processing in classifying the presence and type of cancer. It releases a clear, concise prediction about whether there is cancerous tissue in the scanned lung or not, possibly as specific as providing information on cancer type and stage.

As shown in Fig. 5, the Receiver Operating Characteristic (ROC) analyses representing the diagnostic performance of the proposed CanNS framework in two different configurations. These ROC curves contrast the different threshold settings between true positive rate sensitivity and false positive rate of 1 – specificity, graphically yielding a complete ranking of the models in terms of accuracy. Figure 5(a) illustrates the performance changes when incorporating the SwiNet model into the CanNS framework. The ROC curve with SwiNet reveals how this segmentation model contributes to improved overall performance by locating areas of interest in the lung CT images. In contrast, that without SwiNet is for CanNS in the absence of the segmentation step or even a less effective method. In this case, it would be the area under the receiver operating characteristic curve; the higher it is, with SwiNet incorporation, the better sensitivity and specificity would be, thereby indicating that the model has a better ability to discriminate between cancerous and non-cancerous tissues correctly. Thus, this ROC analysis indeed identifies the contribution of SwiNet as most critical in boosting the diagnosis accuracy of the CanNS framework through high-quality and fine-grained segmentation. Figure 5(b) shows the performance benchmark of the CanNS framework with and without incorporating the Devilish Levy Optimization technique, DevLO, in choosing optimally the learning rates of the XLG CancerNet model for comparatively improved training and predictive performance of the model. The ROC curve that includes DevLO defines a better AUC, thus indicating that this optimization technique contributes much to the discriminative powers of the model to really detect the cancerous versus non-cancerous samples. Contrary to this, an ROC without DevLO will basically return a poorer AUC for this model without such fine-tuning. This analysis has served to further underline the important role DevLO has played in fine-tuning learning parameters to guarantee better generalization and accuracy of the model towards lung cancer diagnosis using CT images.

Figure 6 captures the progress in this context of learning rate tuning for the CanNS framework. The x-axis is the iteration count—the number of iterations the algorithm has run. The y-axis is the fitness value, which itself, as an outcome, holds the measure of performance of that current solution. First, all fitness values are changing

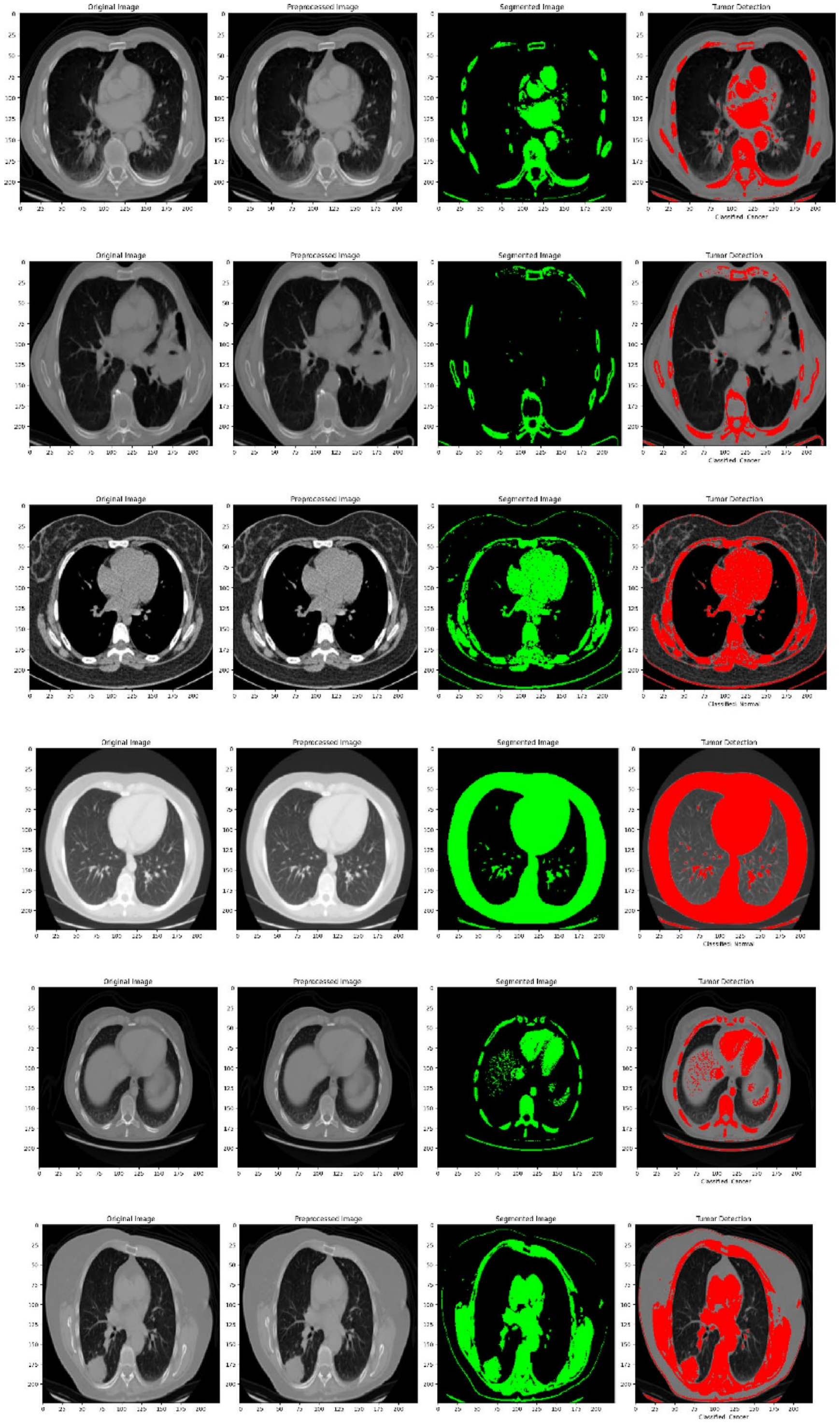
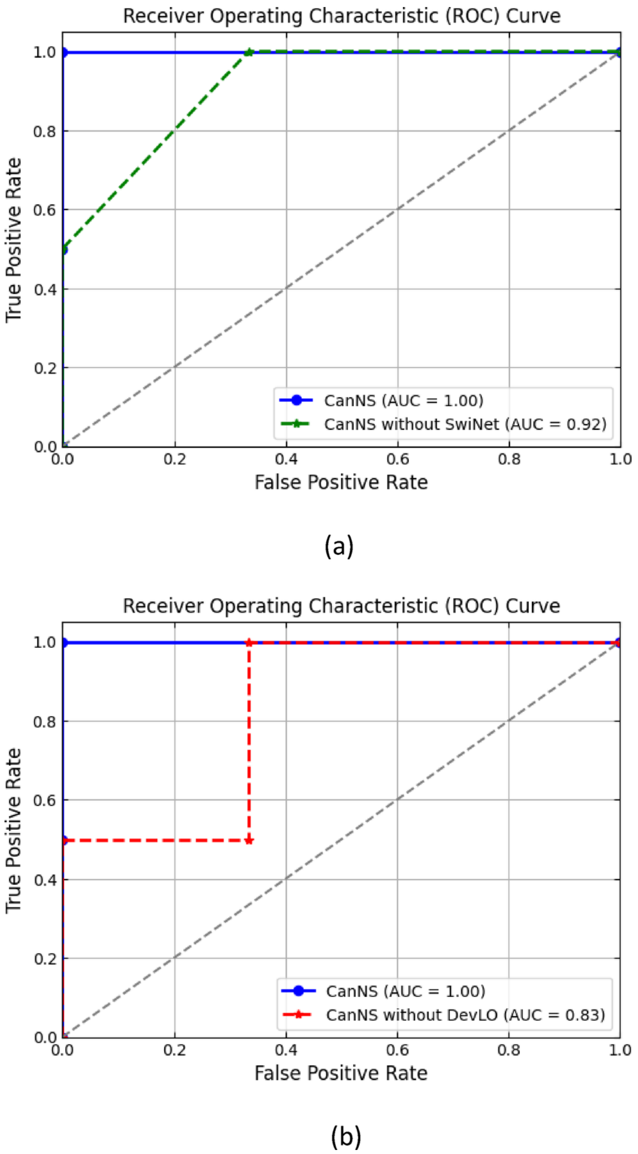


**Fig. 4**. Input CT lung image, preprocessed image, segmented image and final predicted result.

due to the initialization of the algorithm in going through possible solutions, but as iterations proceed, the curve will start to stabilize, hence indicating that it is converging into an optimal solution. This clearly showed the efficacy of the algorithm in arriving at a learning rate that would enhance model performance during this phase of stabilization. The fitness curve obtained in DevLO clearly denotes iterative improvement and fine-tuning of the learning rate, which will definitely go ahead to produce better results while training models.

Figure 7 shows the comparison of accuracy, sensitivity, and specificity of the CanNS model alone, with and without the SwiNet algorithm. This would provide a clear view of how the SwiNet algorithm affects the performance of the CanNS model. Accuracy measures the whole correctness of the model and shows how often it correctly classifies cancerous and non-cancerous images. Sensitivity, otherwise referred to as recall, refers



(a)

(b)

**Fig. 5**. (**a**). ROC analysis of CanNS model with and without SwiNet. (**b**). ROC analysis of CanNS model with and without DevLO technique.

to the percentage of the correctly identified true positive instances of cancer, while specificity refers to the proportion of correctly identified true negatives, meaning non-cancer cases. We can renegotiate that adding the SwiNet algorithm does increase in reality, performance in all three metrics for the CanNS model. With SwiNet, the model will be more accurate and thus make fewer total classification errors. That is to say, with increased sensitivity, the model becomes very powerful in detecting actual cases of cancer in order to reduce the risk of missing any critical diagnosis. At the same time, increased specificity means improved performance in rightly identifying non-cancerous cases, which is critical in avoiding unwarranted treatment or anxiety for the patients. As done with the CanNS model, the evaluation will be repeated using DevLO and without the DevLO technique in Fig. 8. In other words, DevLO will be primarily drafted to optimize the learning rate during the model training phase. Results show that the model-based CanNS with DevLO outperforms the model without according to all three performance criteria: accuracy, sensitivity, and specificity. That means including DevLO in a model will increase the accuracy, which is proof that the optimized learning rate really works on the model to make more effective computations toward making accurate predictions. Improvement in sensitivity means it will be better at picking up cases of cancer—integral to creating an early and accurate diagnosis. It simply means that improved specificity will help the model improve in non-cancer cases identification, hence reducing the possibilities of false positives. These figures therefore underscore the high benefits of integrating the SwiNet algorithm with the DevLO technique into the CanNS framework for developing a more accurate model for the diagnosis of this cancer. Besides these critical metrics, it can be labeled how effective these enhancements were in refining predictive abilities and overall model performance for medical image analyses.

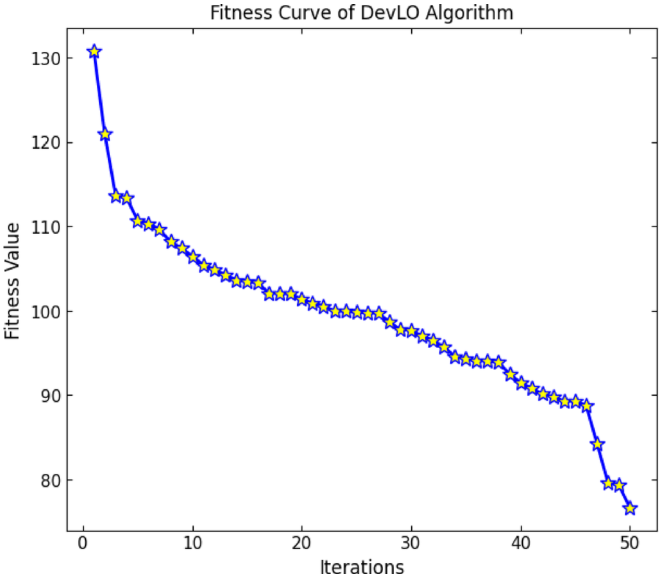


**Fig. 6**. Fitness curve of DevLO algorithm.

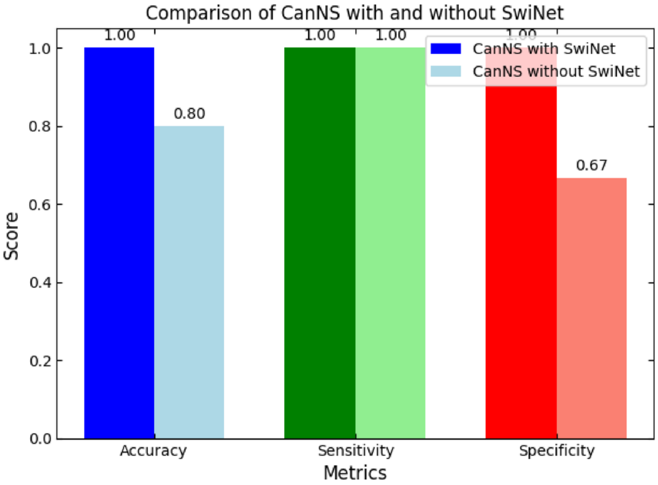


**Fig. 7**. Accuracy, sensitivity and specificity analysis of CanNS model with and without SwiNet algorithm.

Figure 9 demonstrates the training and validation loss of the CanNS with and without the application of the SwiNet. It is important for evaluating the effect of SwiNet on the learning process and model's ability to generalize. The training accuracy is the ratio of the correct predictions made by the model in the training data set while validation accuracy is the ratio of the diffident predictions made by the model in the validation data set but generalize to unseen data. By analyzing the resulted figure, we clearly understand that when we include SwiNet architecture within the CanNS model, the training and validation accuracy are apparently higher as compared with the case when we do not use SwiNet architecture.

This implies that SwiNet is able to improve the model capacity in learning the relevant features of the input data especially when it comes to training as well as validation datasets. The overall, increased, and specific fact is that the accuracy of validation in this case, suggests that the model with SwiNet learns to generalize and perform better on new unseen data, which for medical image diagnosis is crucial. The values of training and validation loss of CanNS model with the application and without applying the SwiNet algorithm are depicted in the Fig. 10. The training loss calculates the difference between the predicted values by the model and what is expected on the training data and the same with a validation loss, which calculates the same thing but this time on the validation data. The loss function implies that lower loss values are desirable since they imply that the model's predictions are closer to the true outcomes. Analyzing the given graph, it can be inferred that by integrating SwiNet, the CanNS model has lesser training and validation loss than the model without SwiNet. This analysis indicates that the SwiNet aids in reducing the errors that the model makes during training and enhances performance when used with new inputs. The lower validation loss is an aspect that stands out as it means that when using SwiNet,

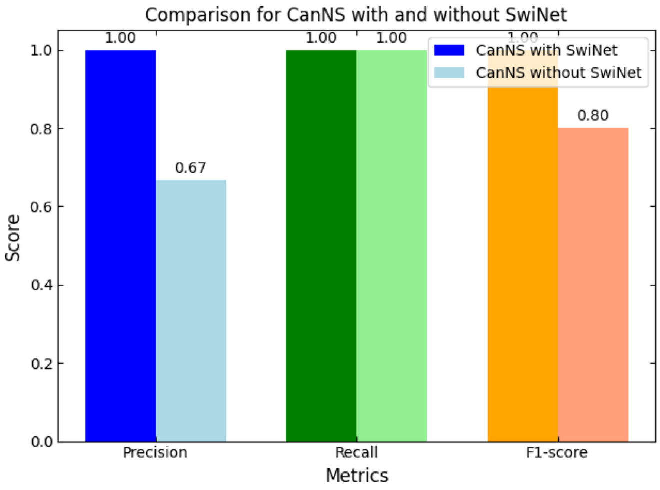


**Fig. 8**. Accuracy, sensitivity and specificity analysis of CanNS model with and without SwiNet algorithm.

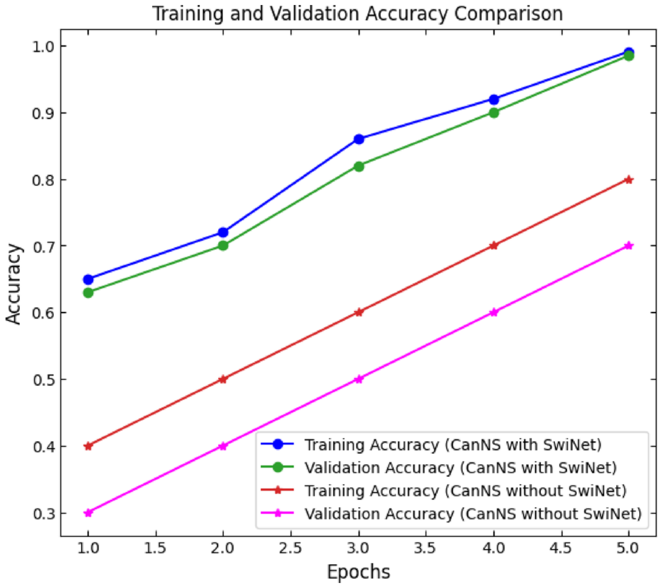


**Fig. 9**. Training and validation accuracy with and without SwiNet.

the model is less prone to over-fitting the training data and is more capable of producing good results on real medical images.

The required training and validation time of the CanNS model is depicted in Fig. 11. This is an important parameter to consider that will help determine how effective the CanNS model is in its working through computational power to train as well as validate the model. Training time is defined as the overall time it takes to train the model on the given dataset whereas validation time is defined as the amount of time taken to validate model on the validation set. As seen from the analysis, it can be observed that the training and Validation time is quite reasonable of the order which is acceptable when working the field of deep learning especially when dealing with medical image data. The training time is higher than the validation time as it would be expected to take longer since it includes processes of optimization to make adjustments on the model's parameters. The validation time, shorter than the training time states that once the model has been trained, the model is capable of faster testing of the data. This efficiency is particularly valued in cases when it is critical to determine the patient's condition as quickly as possible. The results of overall comparative analysis of CanNS model with the recent existing models are shown in Fig. 12. This comparison is based on factors of performance such as accuracy, sensitivity and specificity either in increased or decrease. The figure gives a clear impression on how the presented CanNS model compares to other new state-of-art models in the field of medical image analysis. An analysis of the presented CanNS model shows the effectiveness of the proposed model in comparison to the existing ones in all aspects.

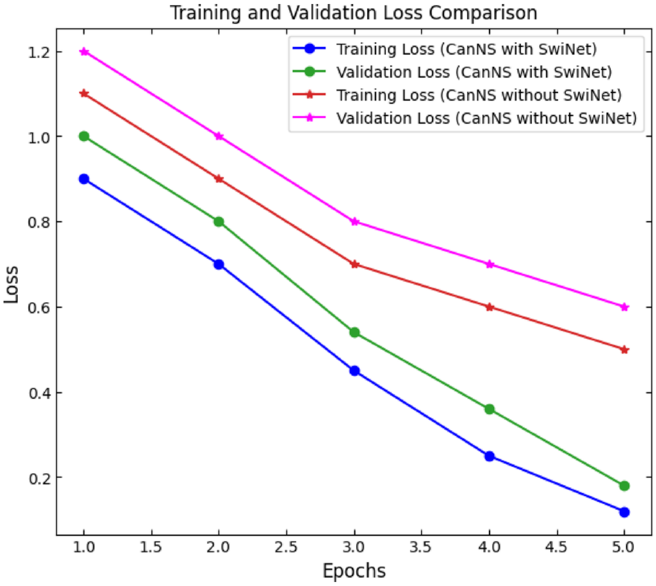


**Fig. 10**. Training and validation loss with and without SwiNet.

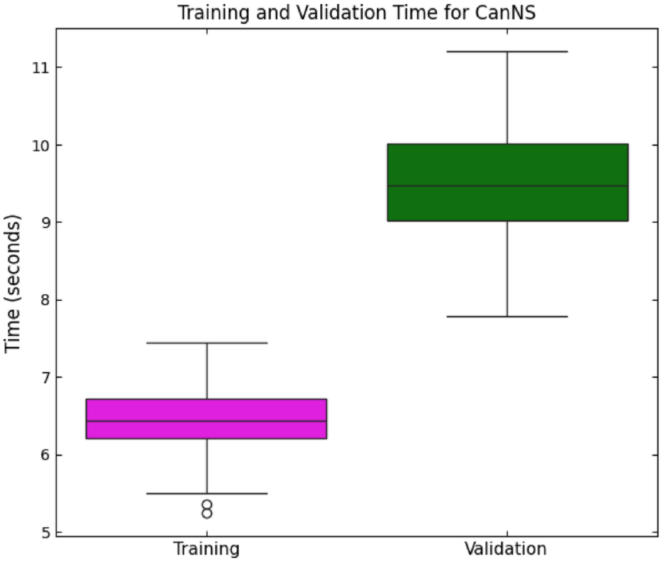


**Fig. 11**. Training and validation time of CanNS model.

The accuracy metric demonstrates that there is a significantly increased ability of CanNS to classify the actual malignant and benign tumors and cases without this condition. The result of the sensitivity metric reveals that CanNS performs well with a higher true positive: this suggests that CanNS is best suited when it comes to identification of cancerous cases. Along the same line, it can be visualized from the specificity metric that CanNS has higher capability of TN that depicts its capacity to identify normal samples where cancer cannot be identified. In comparing the two models, the versatility and effectiveness of CanNS model are well demonstrated. Looking at all the evaluated parameters it will clearly show how SwiNet is more efficient when used in segmentation, XLG in feature extraction and classification as well as how DevLO is efficient in tuning the parameters. Each of the methodologies used in the Can NS model enhances the ability of the model thus making it to be among the best for cancer diagnosis using medical images.

The AUC is another performance measure used in determining how well the model separates the classes; in this case, between malignant and benign lung CT images as shown in Fig. 13. The AUC value vary between 0 and 1, the higher the value the better the performance. From the analysis, it can be inferred that the CanNS model has a higher AUC than this latest existing models; thus, the CanNS model is better in discrimination. This high AUC value portrays that of the CanNS model we categorized the cancer instances and the non-cancer control instances more effectively without many false positives or false negatives. SwiNet is applied for feature

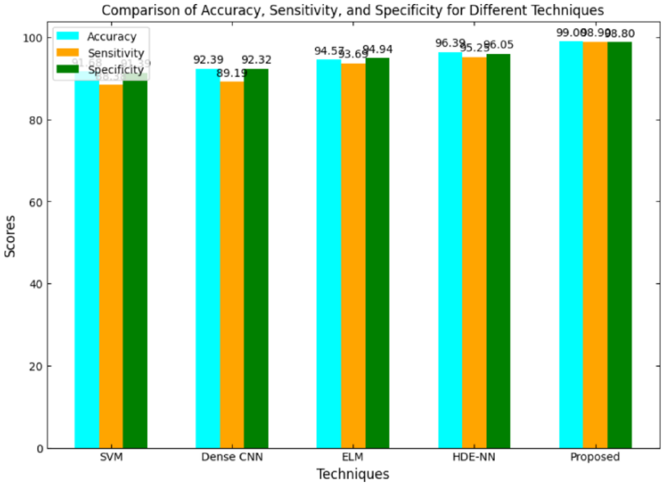


**Fig. 12**. Overall comparative study with recent existing models.

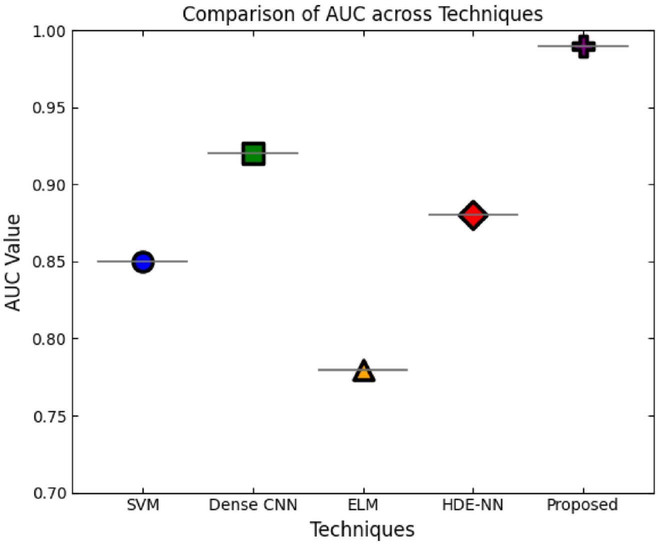


**Fig. 13**. Comparative study based on AUC with recent existing models.

segmentation from the CanNS framework while XLG is used for feature extraction and classification, and lastly, DevLO for parameter optimization, all of which improve the model efficiency.

Segmentation results among current and proposed transformer models is shown in the Fig. 14. Something as simple as segmentation accuracy is crucial in medical image analysis since it quantifies the model's capacity to segment the regions of interest, such as cancerous tissues in lung CT images[41]. It is evident from the figure that the CanNS framework with the proposed transformer model has attained better segmentation accuracy than other transformer models. This improvement can be due to the architecture of SwiNet which is more optimized especially for image segmentation tasks. The transformer models stand out in extracting local intricate patterns and details in the images, which contributes to the improvement of the segmentation results, as they outperform the encoder-decoder architectures. As seen from the comparative analysis, there is an opportunity to implement and use the above proposed transformer model within the CanNS framework. That is the reason a higher segmentation accuracy of the CanNS model presented, as it is proved to be more suitable for the precise identification of cancerous areas in lung CT images, making it more beneficial for doctors in diagnosing cancer patients. It also supports the reliability and speed in the implemented CanNS framework by showing the superiority of other transformer models in medical image segmentation tasks.

In Fig. 15, there is a comparison of relative performances of the conventional classification methodology in addition to the proposed classification method in terms of accuracy, sensitivity, and specificity and significance of the new and CanNS model. These metrics are particularly significant when evaluating classification models particularly in the health field where a wrong diagnosis may be costly. The formula for accuracy is the number of true results which are first the true positives and the second true negatives divided by the total number of cases that have been analyzed. The accuracy parameter represents the average probability of a correct classification

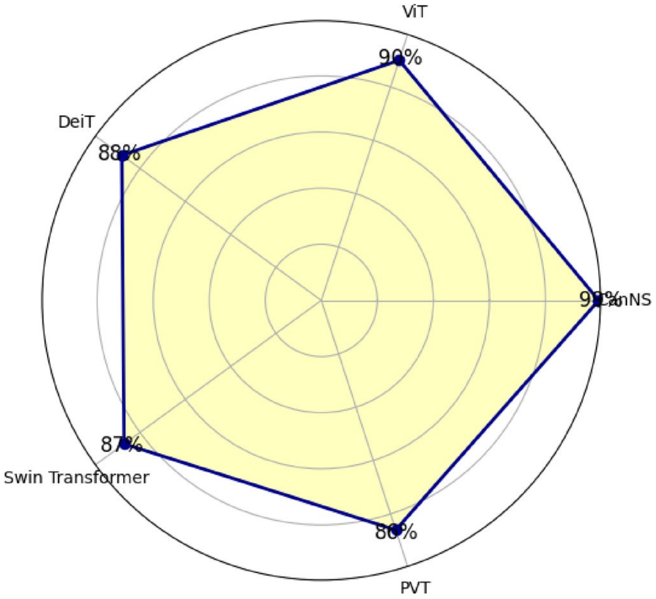


Fig. 14. Comparison among existing and proposed transformer models based on segmentation accuracy.

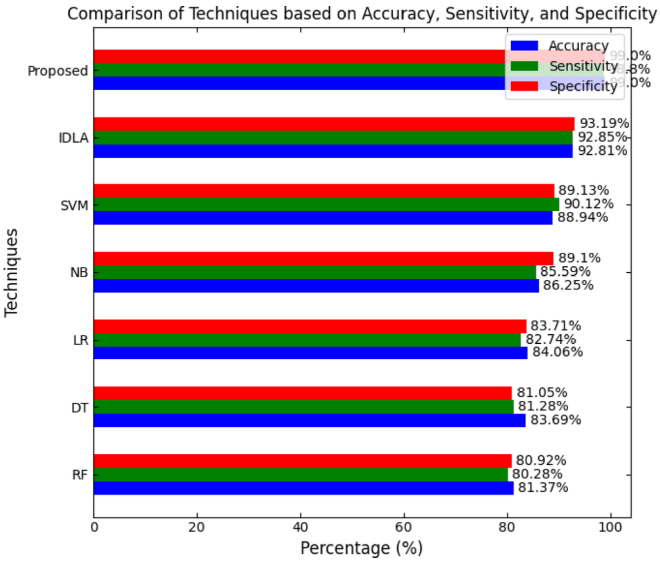


Fig. 15. Comparison among conventional and proposed classification techniques based on accuracy, sensitivity and specificity.

between carcinogenic and non-carcinogenic ones. It means that by having a higher value the model is less likely to make a mistake when diagnosing and therefore the diagnosed is good. Sensitivity, also referred to as recall rate, focuses on the model's capability to accurately identify true positives or the cancerous cases. Consequently, high sensitivity is desirable in the medical diagnostics to achieve a statement that a majority of the cases of cancer disease are discovered in their early stage, which should enhance the prospects of cure of the patients. Accuracy is the measure of the model which regards non-cancerous incidences or true negative correctly identified by the model.

Another essential characteristic, high specificity is also crucial to avoid false positive results and exclude apparently healthy persons from the oncological disease, which can lead to adverse psychological and physical effects, additional examinations, and other injurious actions. The above figure unambiguously reveals the superiority of the CanNS model over conventional classification approaches in all the considered measures. All these factors can be attributed to the integrated nature of CanNS where SwiNet is used to perform accurate segmentation of medical images, XLG is used to extract features and classification of the images and DevLO is used for the optimal tuning of the parameters. Higher accuracy, sensitivity, and specificity of CanNS suggest that the model is capable of diagnosing lung cancer with higher reliability and accuracy compared to the existing

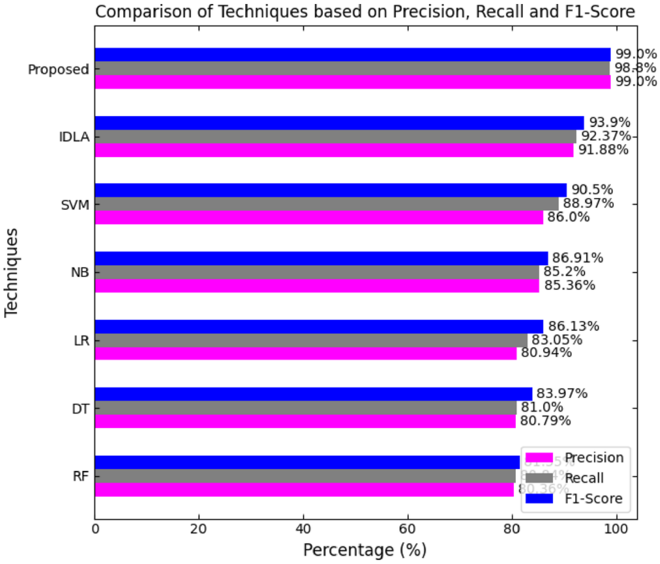


**Fig. 16**. Comparison among conventional and proposed classification techniques based on precision, recall and f1-score.

| Methods | Precision (%) | Recall (%) | F1-score (%) | AUC (%) |
|---|---|---|---|---|
| U-Net with ResNet | 88 | 84 | 86 | 90 |
| VGG16 | 85 | 83 | 84 | 91 |
| DenseNet121 with LSTM | 89 | 87 | 88 | 92 |
| Inception V3 with GAN | 86 | 84 | 85 | 89 |
| Proposed | 98.9 | 98.85 | 98.99 | 99 |

**Table 2**. Comparison with recent deep architecture models.

strategies. Understanding that in every cancer diagnostic there is a comparison between sensitiveness and specifity also guarantee that the CanNS model, apart from identifying higher number of cancer cases in its true sense, is not likely to misdiagnose more healthy people as sick thus making the CanNS model a more reliable and balanced diagnostic tool.

The precision score, recall, and F1-score comparisons of conventional and proposed classification techniques are depicted in Fig. 16. These metrics are very useful when considering the detailed performance of classification models. Precision defines the number of correct positives over the total number of positive predictions, which shows the ability of the model not to include healthy people in the cancer group. High precision means a lower false positive rate, an important factor in a health management setup to avoid costly and time wasting procedures which could bring unnecessary discomfort to patients. Sensitivity, or recall, quantifies the logical correctness of the model as concerning true positives. This is important in attaining the objective of identifying most of the cancer cases especially at early stage since treatment is most effective. The f1-score is the average of the accuracy (precision) and the completeness (recall) scores which allows to consider at the same time the number of false positive and false negative cases. This metric is especially effective if the distribution of classes is skewed because it provides a more detailed assessment of the model's performance. What can be observed in the above figure is that the under consideration CanNS model attains a high level of precision, recall, and f1-score more than the classification model. This enhancement is achieved through the combination of; SwiNet for efficient segmentation function, XLG for feature extraction and classifying functions, and DevLO for optimizing the learning parameters. Higher accuracy indicates that the model misdiagnoses fewer cancer cases, and higher recall means a greater proportion of the actual cancer cases is found, F1-score gives a balanced result. Overall, the presented CanNS model demonstrates enhanced accuracy, reliability, and encompassing nature that are significant in the cancer diagnosis, which makes it an effective tool in the sphere of medicine. In Table 2, comparison of the proposed CanNS Framework with other state-of-the-art deep learning models concerning the detection performance for lung cancer using CT scan images. It can be observed from the table that, among all metrics, the proposed CanNS Framework with Swin-Transformer UNet segmentation, Xception-LSTM GAN CancerNet classification, and Devilish Levy Optimization for fine-tuning parameters has the best performance- it can achieve a Precision of 98.9%, a Recall of 98.85%, an F1-score of 98.99%, and an AUC of 99% performance. It depicts that the model is with great potential ability for the diagnosis of lung cancer with as low as possible false positive and negative values. Comparatively lesser performances were in U-Net combined ResNet, VGG16, DenseNet121-LSTM, InceptionV3 with GAN. U-Net with ResNet depicts a precision of 88%, an recall of 84%, F1-score of 86%, and AUC of 90%, while VGG16 reports results with 85% in precision, 83% recall, and 91% in AUC. The DenseNet121 combined with LSTM gives better results in terms of 89% precision, 87% recall, and 92% AUC but still lags far behind the CanNS Framework. Among these, the inception V3 with GAN slightly outperformed others like VGG16 and U-Net with ResNet but failed to outperform the CanNS Framework, which has achieved 86% precision, 84% recall, and 89% AUC. These comparisons indeed prove that the performance of detection in the medical images for lung cancer is enhanced significantly with good accuracy, sensitivity, and specificity by the CanNS Framework; hence, it is efficient and reliable.

Among these, adapting the CanNS model to a clinical environment in real-world situations involves an essential challenge given that there are differences in the kind of imaging protocols available. Some influencing factors for good or poor scanning in CT include types of scanners, slice thickness, contrasting, resolution, and exposure settings, all of which would make much difference between two scans. In order to be sure of this, the model needs the CanNS to integrate several powerful preprocessing techniques to standardize the images from different scanners with diverse protocols. One potential path for developing an adaptive pipeline using techniques such as histogram equalization, rescaling, and noise reduction so as to minimize the variations presented because of protocols.

These domain adaptation techniques may be considered in parallel with an attempt to fine-tune the model for a given clinical setting or scanner. Due to the multimodal approach, which may be supported with different segmentation and classification and optimization components, the architecture of the model will also be flexible. This means that, for clinical deployment, constant monitoring and updating will be necessary so model performance maintains its accuracy and reliability concerning new imaging protocols and patient populations. Finally, an intuitive interface must be provided to clinicians that can manage such variations and provide interpretability in the decisions of the model will be crucial in gaining trust and ensuring CanNS is useful and safe in clinical practice.

## Conclusion

This work presents the CanNS framework which entails a broad strategy of improving the effectiveness of identifying lung cancer by leveraging on enhanced machine learning heuristics. The CanNS framework uses the SwiNet segmentation model and the XLG CancerNet classification model and DevLO algorithm for parameter tuning that has its own advantage in the system. SwiNet, used for the current study in the segmentation part, shows a high capability in differentiating lung nodules from CT pictures. The architecture which combines both convolutional and attention mechanisms makes the segmentation precise in the sense that both local as well as global features are captured. The delineation of such features is important since they help in determining early stage lung cancer, and the result is improved visualization of such tissues. For the classification of the segmented regions, XLG CancerNet is used. This model takes advantage of transformer-based or transformer-like networks to achieve good performance in differentiating between malignant and benign nodules. XLG CancerNet has self-attention for transformers that allow the consideration of the entire context of an image, which is not noted in CNN and improves its performance in the detection of cancerous patterns. This leads to improved and more stable classification results which are critical for diagnosis. The DevLO algorithm works as a key component in tuning all the parameters of the proposed CanNS arrangement.

Thus, DevLO successfully optimizes the value of other hyperparameters so that both SwiNet as well as XLG CancerNet work most efficiently. This optimisation makes the models to be more general and this increase the validity of the model, in the different datasets. As mentioned earlier, DevLO is employed to strike the best balance for time computational efficiency and models, thus makes CanNS practical for clinical based applications. The results of this work are the CanNS framework's validation on the Kaggle lung CT image dataset, proving its effectiveness compared to other similar methods. It was found that the accuracies of detection are much higher; also the sensitivity and specificity and thus the aforementioned framework can be vital in early and correct diagnosis of lung cancer. These advancements can counter some of the weaknesses of prior models; large annotated data for training, overfitting and high-computational costs, to name but a few. Altogether, the CanNS for the lung cancer detection framework consisting of integration of SwiNet for boundary segmentation, XLG CancerNet for classification, and DevLO for parameter optimization, large-scale lung cancer detection has advanced remarkably significantly. The given work raises the hope for the application of this enhanced framework in the clinical practice and could bring beneficial results in terms of patients' outcomes due to accurate and efficient diagnosis. The limitation could be that the whole research had to rely on only one kind of imaging protocol, which could imply limited variety of the CT scan data and make generalizations for other types of imaging devices or clinical setups limited. Secondly, the data is imbalanced since the number of benign cases is fewer compared to the rest. In the future research, there will be attempts to make the further adjustment of the proposed framework, as well as its use will be discussed to be applied not only to the mentioned types of cancer but also to other types of cancer and different types of medical imaging techniques in order to bring the further benefit to the field and practice of medical diagnostics. Immediately within the near future, the research in the domain of lung cancer detection by means of deep learning frameworks similar to CanNS will concentrate on an increase in model generalization at large-scale and diverse datasets resulting from multiple clinical settings. The first is model improvement regarding robustness, such as considering variations in quality within the images, scanner type, and patient demographics-features that highly impact models' performance. Besides, the researchers will probe into ways of model interpretability since the clinical usage of AI in healthcare can afford nothing less than crystal-clear decision-making. The main challenges are: computational constraints, real-time performance in the clinical setting, and keeping the accuracy high. In a future prospect, the system could be further developed in an autonomous learning way, with continuous improvement, case by case, enriched by feedback from real usage, therefore continuously improving diagnostic performance over time.

## Data availability

The data that support the findings of this study are available from the corresponding author, upon reasonable request.

Received: 22 August 2024; Accepted: 29 April 2025

## References

1. Hosseini, S. H., Monsefi, R. & Shadroo, S. Deep learning applications for lung cancer diagnosis: a systematic review. *Multimedia Tools Appl.* **83**, 14305–14335 (2024).
2. BR, S. R. R. et al. Stacked neural Nets for increased accuracy on classification on lung cancer. *Measurement: Sens.* **32**, 101052 (2024).
3. Chung, J. H., Park, J. M. & Kim, D. H. Automated CT quantification of interstitial lung abnormality in patients with resectable stage I non-small cell lung cancer: prognostic significance. *Thorac. Cancer*, **15**(16), 1305–1311 (2024).
4. Subash, J. & Kalaivani, S. Dual-stage classification for lung cancer detection and staging using hybrid deep learning techniques. *Neural Comput. Appl.* **36**, 8141–8161 (2024).
5. Meng, L., Zhu, P. & Xia, K. Application value of the automated machine learning model based on modified CT index combined with serological indices in the early prediction of lung cancer. *Front. Public. Health.* **12**, 1368217 (2024).
6. Lu, M. T., Raghu, V. K., Mayrhofer, T., Aerts, H. J. & Hoffmann, U. Deep learning using chest radiographs to identify high-risk smokers for lung cancer screening computed tomography: development and validation of a prediction model. *Ann. Intern. Med.* **173**, 704–713 (2020).
7. Paramasivam, R., Patil, S. N., Konda, S. & Hemalatha, K. Lung cancer computed tomography image classification using attention based capsule network with dispersed dynamic routing. *Expert Syst.* **41**, p. e13607 (2024).
8. Bhandary, A. et al. Deep-learning framework to detect lung abnormality–A study with chest X-Ray and lung CT scan images. *Pattern Recognit. Lett.* **129**, 271–278 (2020).
9. Quanyang, W. et al. Artificial intelligence in lung cancer screening: detection, classification, prediction, and prognosis. *Cancer Med.* **13**, e7140 (2024).
10. Sivakumar, M., Chinnasamy, S. & Thanabal, M. An efficient combined intelligent system for segmentation and classification of lung cancer computed tomography images. *PeerJ Comput. Sci.* **10**, e1802 (2024).
11. Mahmood, T. et al. A brief survey on breast Cancer diagnostic with deep learning schemes using Multi-Image modalities. *IEEE Access.* **8**, 165779–165809 (2020).
12. Gayap, H. T. & Akhloufi, M. A. Deep machine learning for medical diagnosis, application to lung cancer detection: a review, *BioMedInformatics*. **4**, 236–284 (2024).
13. Barbouchi, K. et al. A transformer-based deep neural network for detection and classification of lung cancer via PET/CT images. *Int. J. Imaging Syst. Technol.* **33**, 1383–1395 (2023).
14. Ali, H., Mohsen, F. & Shah, Z. Improving diagnosis and prognosis of lung cancer using vision Transformers: a scoping review. *BMC Med. Imaging*. **23**, 129 (2023).
15. Liu, Y. et al. Lung nodule malignancy classification with associated pulmonary fibrosis using 3D attention-gated convolutional network with CT scans. *J. Translational Med.* **22**, 51 (2024).
16. Prabakaran, J. & Selvaraj, P. An approach to recognise lung diseases using segmentation and classification. *Meas. Sci. Rev.* **23**, 254–259 (2023).
17. Mkindu, H., Wu, L. & Zhao, Y. Lung nodule detection in chest CT images based on vision transformer network with bayesian optimization. *Biomed. Signal Process. Control.* **85**, 104866 (2023).
18. Mahmood, T., Saba, T., Rehman, A. & Alamri, F. S. Harnessing the power of radiomics and deep learning for improved breast cancer diagnosis with multiparametric breast mammography. *Expert. Syst. Appl.*. **249**, 123747 (2024).
19. Da Nobrega, R. V. M. et al. Lung nodule malignancy classification in chest computed tomography images using transfer learning and convolutional neural networks. *Neural Comput. Appl.* **32**, 11065–11082 (2020).
20. Chen, H. Y., Wang, H. M., Lin, C. H., Yang, R. & Lee, C. C. Lung Cancer prediction using electronic claims records: A Transformer-Based approach. *IEEE J. Biomedical Health Inf.* **27**(12), 6062–6073 (2023).
21. Shashikala, D., Chandran, C. & Rajathi, S. Cross-spectral vision transformer for lung nodule detection with improved moth flame algorithm using deep learning. *e-Prime-Advances in Electrical Engineering, Electronics and Energy*, **8**, 100556 (2024).
22. Yang, S. et al., Extracting Pulmonary Nodules and Nodule Characteristics from Radiology Reports of Lung Cancer Screening Patients Using Transformer Models. *J. Healthc. Inf. Res.* **8**, 463–477 (2024).
23. Mahmood, T. et al. Breast lesions classifications of mammographic images using a deep convolutional neural network-based approach. *PLOS ONE*. **17**, e0263126 (2022).
24. Kunkyab, T., Bahrami, Z., Zhang, H., Liu, Z. & Hyde, D. A deep learning-based framework (Co–ReTr) for auto–segmentation of non–small cell–lung cancer in computed tomography images. *J. Appl. Clin. Med. Phys.* **25**, e14297 (2024).
25. Mahmood, T., Li, J., Pei, Y. & Akhtar, F. An Automated In-Depth Feature Learning Algorithm for Breast Abnormality Prognosis and Robust Characterization from Mammography Images Using Deep Transfer Learning. *Biology*. **10**, 859 (2021).
26. Quasar, S. R., Sharma, R., Mittal, A., Sharma, M. & Agarwal, D. Torre Díez, ensemble methods for computed tomography scan images to improve lung cancer detection and classification. *Multimedia Tools Appl.* **83**, 52867–52897 (2024). de La.
27. Shah, S. N. A. & Parveen, R. An extensive review on lung cancer diagnosis using machine learning techniques on radiological data: state-of-the-art and perspectives. *Arch. Comput. Methods Eng.* **30**, 4917–4930 (2023).
28. Thanoon, M. A., Zulkifley, M. A., Mohd Zainuri, M. A. A. & Abdani, S. R. A review of deep learning techniques for lung cancer screening and diagnosis based on CT images. *Diagnostics*. **13**, 2617 (2023).
29. Siddiqui, E. A., Chaurasia, V. & Shandilya, M. Detection and classification of lung cancer computed tomography images using a novel improved deep belief network with Gabor filters. *Chemometr. Intell. Lab. Syst.* **235**, 104763 (2023).
30. Nigudgi, S. & Bhyri, C. Lung cancer CT image classification using hybrid-SVM transfer learning approach. *Soft. Comput.* **27**, 9845–9859 (2023).
31. Lanjewar, M. G., Panchbhai, K. G. & Charanarur, P. Lung cancer detection from CT scans using modified densenet with feature selection methods and ML classifiers. *Expert Syst. Appl.* **224**, 119961 (2023).
32. AR, B., RS, V. K. & SS, K. LCD-capsule network for the detection and classification of lung cancer on computed tomography images. *Multimedia Tools Appl.* **82**, 37573–37592 (2023).
33. Dunn, B., Pierobon, M. & Wei, Q. Automated classification of lung cancer subtypes using deep learning and CT-scan based radiomic analysis, *Bioengineering*. **10**, 690 (2023).
34. Murthy, S. & Prasad, P. M. K. Adversarial transformer network for classification of lung cancer disease from CT scan images. *Biomed. Signal Process. Control.* **86**, 105327 (2023).
35. Lyu, Q. et al. A transformer-based deep-learning approach for classifying brain metastases into primary organ sites using clinical whole-brain MRI images. *Patterns*. **3**, 100613 (2022).

36. Xiao, H., Li, L., Liu, Q., Zhu, X. & Zhang, Q. Transformers in medical image segmentation: A review. *Biomed. Signal Process. Control.* **84**, 104791 (2023).
37. Mahmood, T., Rehman, A., Saba, T., Nadeem, L. & Bahaj, S. A. O. Recent advancements and future prospects in active deep learning for medical image segmentation and classification. *IEEE Access.* **11**, 113623–113652 (2023).
38. Arshad, W. et al. Cancer unveiled: A deep dive into breast tumor detection using Cutting-Edge deep learning models. *IEEE Access.* **11**, 133804–133824 (2023).
39. Asuntha, A. & Srinivasan, A. Deep learning for lung Cancer detection and classification. *Multimedia Tools Appl.* **79**, 7731–7762 (2020).
40. Gugulothu, V. K. & Balaji, S. An early prediction and classification of lung nodule diagnosis on CT images based on hybrid deep learning techniques. *Multimedia Tools Appl.* **83**, 1041–1061 (2024).
41. Fanizzi, A. et al. Comparison between vision Transformers and convolutional neural networks to predict non-small lung cancer recurrence. *Sci. Rep.* **13**, 20605 (2023).

## Author contributions

All authors have equally contributed to this manuscript. All authors reviewed the manuscript.

## Funding

No funding was supported for this research work.

## Declarations

## Competing interests

The authors declare no competing interests.

## Ethics approval

No human subject or animals are involved in the research.

## Additional information

**Correspondence** and requests for materials should be addressed to S.S.

**Reprints and permissions information** is available at www.nature.com/reprints.