CAAI Transactions on Intelligence Technology

WILEY

The Institution of Engineering and Technology

ORIGINAL RESEARCH OPEN ACCESS

VSMI^2-PANet: Versatile Scale-Malleable Image Integration and Patch Wise Attention Network With Transformer for Lung Tumour Segmentation Using Multi-Modal Imaging Techniques

Nayef Alqahtani[1] | Arfat Ahmad Khan[2] | Rakesh Kumar Mahendran[3] | Muhammad Faheem[4,5]

[1]Department of Electrical Engineering, College of Engineering, King Faisal University, Al-Ahsa, Saudi Arabia | [2]Department of Computer Science, College of Computing, Khon Kaen University, Khon Kaen, Thailand | [3]Department of Computer Science and Engineering, Saveetha Engineering College, Chennai, India | [4]School of Technology and Innovations, University of Vaasa, Vaasa, Finland | [5]VTT Technical Research Centre of Finland Ltd., Espoo, Finland

Correspondence: Muhammad Faheem (muhammad.faheem@uwasa.fi)

Received: 13 December 2023 | **Revised:** 3 March 2025 | **Accepted:** 9 April 2025

Funding: The work of Muhammad Faheem is supported by the VTT Technical Research Centre of Finland and the work of Nayef Alqahtani is supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia (Grant KFU251882).

Keywords: computational intelligence | computer vision | data fusion | deep learning | feature extraction | image segmentation

ABSTRACT

Lung cancer (LC) is a major cancer which accounts for higher mortality rates worldwide. Doctors utilise many imaging modalities for identifying lung tumours and their severity in earlier stages. Nowadays, machine learning (ML) and deep learning (DL) methodologies are utilised for the robust detection and prediction of lung tumours. Recently, multi modal imaging emerged as a robust technique for lung tumour detection by combining various imaging features. To cope with that, we propose a novel multi modal imaging technique named versatile scale malleable image integration and patch wise attention network (VSMI^2 – PANet) which adopts three imaging modalities named computed tomography (CT), magnetic resonance imaging (MRI) and single photon emission computed tomography (SPECT). The designed model accepts input from CT and MRI images and passes it to the VSMI^2 module that is composed of three sub-modules named image cropping module, scale malleable convolution layer (SMCL) and PANet module. CT and MRI images are subjected to image cropping module in a parallel manner to crop the meaningful image patches and provide them to the SMCL module. The SMCL module is composed of adaptive convolutional layers that investigate those patches in a parallel manner by preserving the spatial information. The output from the SMCL is then fused and provided to the PANet module. The PANet module examines the fused patches by analysing its height, width and channels of the image patch. As a result, it provides an output as high-resolution spatial attention maps indicating the location of suspicious tumours. The high-resolution spatial attention maps are then provided as an input to the backbone module which uses light wave transformer (LWT) for segmenting the lung tumours into three classes, such as normal, benign and malignant. In addition, the LWT also accepts SPECT image as input for capturing the variations precisely to segment the lung tumours. The performance of the proposed model is validated using several performance metrics, such as accuracy, precision, recall, *F*1-score and AUC curve, and the results show that the proposed work outperforms the existing approaches.

1 | Introduction

Cancer is one of the most dangerous and common diseases affecting people worldwide. It is centred on the spread and multiplication of abnormal cells within the human body. The early diagnosis and timely detection provide a promising path to efficient therapy [1]. As a part of normal biological cycles, damaged and ageing cells are automatically replaced by new ones. When this process breaks down and injured cells survive, the cancer develops uncontrollably. These abnormal cells can potentially spread to other organs in the body [2].

Lung cancer, which accounts for 15.7% of all cancers diagnosed worldwide, is a well-known cause of cancer-related fatalities, and it develops in lung tissue. When discovered at an early stage, there is a 79% likelihood of a 5-year survival rate compared to only 10% at an advanced stage [3]. The early diagnosis and treatment can also dramatically increase survival chances. Unfortunately, an early-stage lung cancer frequently goes undetected, and it becomes symptomatic later on [4]. The 5-year survival rate in the United States of America is only 20.7%, and in developing nations, it is significantly lower. The early detection not only speeds up recovery but also makes therapy easier and less expensive, greatly improving the likelihood of recovery. Early detection and treatment in the USA can increase the 5-year survival rate from 21% [5].

The National Lung Screening Trial (NLST), which attempted to reduce lung cancer related mortalities by 23%, demonstrated that the early identification enhances the hopes of high-risk patients [6]. The main obstacle in screening techniques is related to the identification of lung nodules. The nodule detection rates are adversely impacted by several elements such as radiologist tiredness, higher workloads and severe turn-around time requirements [7]. Numerous studies have emphasised the prevalence of diagnostic errors in clinical practice which can be attributed to a variety of factors including patient-specific, nodule-specific (i.e., size and density) and environment-specific (i.e., lack of equipment and shortages and higher workload) problems.

Both computed tomography (CT) and magnetic resonance imaging (MRI) are well-known medical procedures used for early detection, increasing the likelihood of patient survival [8]. However, recently adopted imaging methods, such as positron emission tomography (PET), and single photon emission computed tomography (SPECT), are also widely adopted for lung cancer diagnosis [9]. These methods are used by radiologists to examine and evaluate whether any troublesome lung nodules are presented or not. Radiologists frequently have to look through a large number of slices from a single patient's image dataset taken from CT, MRI, SPECT or PET scans [10]. Radiologists may experience severe fatigue as a result of this manual approach, which increases the risk of errors, such as false positives (identification of normal tissue as a nodule) or false negatives (missing nodule detections), in the diagnosis of pulmonary nodules. In order to automate the identification of lung nodules in the collected scan pictures, computer-aided detection (CAD) methods were developed [11].

Most CAD systems that are used to diagnose pulmonary nodules adopt a two-phase strategy which includes the identification of possible nodules and the lowering of false positive rates [12]. The CAD system scans the patient's CT pictures in the early phase to find as many potential lung nodule candidates as it can, essentially locating lung nodules. Although this phase often achieves a high level of sensitivity, it also frequently results in a sizeable proportion of false positive lung nodule identifications [13–15]. As a result, in the second phase, the binary classifier evaluates the identified prospective nodule candidates with the goal of reducing the incidence of false positives in the diagnosis of pulmonary nodules. Conventionally, many of the research works utilised handmade features for lung tumour detection which includes intensity and morphological features, respectively [16]. State-of-the-art classifiers are utilised for processing the handmade features such as support vector machine (SVM), light gradient boosting machine (LGBM), K-nearest neighbours (KNN) etc., However, those adopted models suffer from poor accuracy and complexity issues because of the adoption of handmade features. In addition to that, the major feature distinguishment also increases the false positive rates [17].

Deep learning (DL) algorithms named convolutional neural network (CNN), deep convolutional neural network (DCNN) and several convolutional based models are utilised for analysing the features and variations associated with the lung nodule classification [18]. To be more precise, when the images are provided as an input to the DL models, then these models automatically extract their hidden features with deeper analysis to magnify the difference among the tumour's distribution [19, 20]. Deep learning (DL)-based models have achieved great success in medical picture segmentation; however, they have numerous drawbacks. One of the main limitations is that they rely on enormous datasets. Medical imaging datasets are frequently limited in size because of privacy issues, expensive annotation costs and heterogeneity in clinical settings, making it difficult to train deep models efficiently. This frequently leads to overfitting, in which the model performs well on training data but fails to generalise to new circumstances. Another disadvantage is the high processing cost associated with deep models. CNNs and transformers are examples of deep learning architectures that need substantial processing power and memory, particularly when dealing with high-resolution medical imaging like CT, MRI and SPECT scans. Furthermore, most DL models struggle to capture multi-modal data because combining information from many imaging modalities (e.g., CT, MRI, SPECT) necessitates advanced approaches. These models frequently focus on a single modality, limiting their ability to completely include complementary anatomical and functional information. Furthermore, DL-based segmentation algorithms may be susceptible to differences in image quality, noise and artefacts found in real-world clinical data.

To this end, this research proposes a novel lung tumour segmentation model using malleable convolutional attention layers and lite weight transformer models, respectively, from multi modal images such as CT, MRI and SPECT. The designed model examines the performance of the proposed model with state-of-the-art research works and models. Although our suggested approach is intended for multi-modal fusion, it is critical to assess each modality (MRI and SPECT) separately in order to understand how they contribute to tumour segmentation. MRI

and SPECT images provide complimentary forms of information, MRI for deep structural views and SPECT for functional insights, making it useful to examine their performance independently before demonstrating the benefits of combining them. In addition to that, the benefits of independent analysis based on the modality specific strengths, performance benchmarking and multi modal integration foundation respectively. The major reason for designing the novel model is because we tend to show the variations of lung features to provide the most accurate and reliable tumour segmentation.

The main novelties of this paper are summarised as follows:

- *Multi modal integration for tumour segmentation:* The suggested technique uses a versatile scale-malleable image integration (VSMI^2) module to incorporate complimentary information from CT, MRI and SPECT images, enhancing the accuracy of lung tumour segmentation by exploiting both anatomical and functional data.
- *Patch wise attention network (PANet) with light wave transformer (LWT):* The research describes a unique patch-wise attention network (PANet) combined with a light wave transformer (LWT) that improves feature extraction and context sensitivity at different scales, capturing tiny features in tumour borders for more precise segmentation.

Also, the proposed methodology significantly reduces the issues in the exiting work as follows:

- *Limited precision in tumour boundary detection:* With the single modality images, segmenting lung tumour is highly difficult because of its complex size and shape of tumour. CT images provide effective details of spatial resolution but lack in segmenting the soft tissues, whereas, soft tissue details are provided by the MRI and lack spatial resolution details as the proposed work combines both of the modalities along with the introduction of versatile scale malleable image integration (VSMI^2) module for amplifying the soft tissue differential and spatial resolution details that provides more precise segmentation.
- *Functional information lackness in tumour segmentation:* The conventional works are highly focused on structural imaging but lack in capturing the functional and metabolic aspects leads to poor identification of malignant regions. However, the integration of SPECT adds the functional layer and complementing the structural data from CT and MRI. The proposed work resolves the mentioned challenge by amalgamating SPECT image to the segmentation pipeline leading to comprehensive view of tumour morphology details.
- *Lung tumour scale variability:* As the lung tumours are varied in sizes among modules and masses, it is difficult to capture and process them in varied scales. Many of the conventional models lacked in accurately detecting the tumours in varied scales. The proposed patch wise attention network (PANet) along with light wave transformer (LWT) resolves these issues by integrating scale malleable attention to every image patches. The adaptation of those networks in the proposed model allows us to effective segment the tumours on varied sizes.

2 | Related Works

In Liu et al. [21], the authors propose a federated learning model based on dual path architecture by using ResNet 18 with the aim of detecting lung nodule. However, the model faces issues in terms of generalisability and scalability. In Chen et al. [22], the authors present a pioneering technique for detecting non-small cell lung cancer in 3D PET/CT scans. The approach integrates attention-guided mechanisms to enhance the accuracy of detection, facilitating early diagnosis. In Tortora et al. [23], by utilising the benefits of RadioPathomics, the authors propose a multifaceted approach that integrates diverse data modalities to tailor radiotherapy for non-small cell lung cancer patients. The method aims to optimise treatment based on patient-specific characteristics. The authors in Bian et al. [24] introduce a groundbreaking strategy that combines domain adaptation and zero-shot learning to streamline the annotation process in multi-modality medical image segmentation. This approach minimises the need for extensive manual labelling. The fusion of cross-modality synthesis techniques to improve the lung tumour segmentation in multi-modal MRI images is introduced in Li et al. [25]. By synthesising complementary information from different modalities, this method enhances the segmentation accuracy. An advanced multimodal fusion technique for the classification of lung cancer is proposed by the authors in Barrett and Viana [26]. This method enhances the fusion of various data modalities, leading to improved accuracy in cancer classification. The authors in Caruso et al. [27] present a multimodal ensemble approach designed to predict overall survival in non-small-cell lung cancer patients. This method uses multi-objective optimisation to create a robust prediction model. In Li et al. [28], the authors introduce an innovative deep learning framework that utilises a mask-guided attention mechanism for predicting distant metastasis in lung cancer cases. This approach enhances the accuracy of metastasis prediction. In Syed Musthafa et al. [29], the authors work on proposing a hybrid machine learning algorithm for predicting lung nodules at the early stage from medical images. However, the proposed algorithm depends on the quality of training data which ultimately affect performance. In Balcı et al. [30], the authors propose a series-based deep learning methodology for classifying lung nodules with the aim of increasing diagnostic accuracy. However, the complexity of the model increases due to the series-based architecture. In Suzuki et al. [31], the authors propose a modified 3D U-Net deep learning model for detecting nodule detection in an automatic way. The model faces issues in terms of dependence on specific datasets. In Bhattacharyya et al. [32], the authors work on improving the delineation of nodules by utilising a bi-directional deep learning-based methodology. However, the model faces accuracy issues when images are of low quality. In Gugulothu et al. [33], a hybrid deep learning based algorithm was proposed by the authors with the aim of predicting and classifying lung nodules on CT images. However, because of the hybrid nature, the model faces complexity issues.

In Ruan et al. [34], the authors present the development of a deep learning model for automatic scan range setting in low-dose CT lung cancer screening, streamlining the imaging process. In Ahmed et al. [35], the authors work on detecting and classifying pulmonary nodules using a deep learning-based architecture. However, the approach faces problems in terms of

model's interpretability. In Cui et al. [36], the authors work on a lung nodule detection system by using deep learning. However, the model faces generalizability problems in terms of addressing diverse populations. In Manickavasagam et al. [37], a computer aided detection (CAD) system has been proposed for detecting lung nodule with the help of CNN. Overall, the model works effectively, but it faces challenges in regions where such datasets are scarce because the model is heavily dependent on labelled training data. In Huang et al. [38], the authors explore a self-supervised transfer learning through domain adaptation for classifying benign and malignant lung nodules on thoracic CT scans, improving classification accuracy. The authors in Refs. [39, 40] discuss a cloud-based lung tumour detection and stage classification using deep learning techniques, enabling remote diagnosis and classification of lung tumours. In fact, a deep feature selection and decision-level fusion approach is proposed for lung nodule classification, highlighting the importance of feature selection and fusion techniques in improving the classification outcomes.

3 | Materials and Methods

This research exploits two databases, namely the Harvard database and the LUNA 16 dataset. From the Harvard database, we acquire MRI and SPECT images. The 360 image pairs are collected from the database with a 512 × 512 pixel range. The acquired images are fairly separated into testing and training sets (80:20). The 288 pairs are utilised as training samples, whereas the remaining are testing samples. To further improve the training data, we perform a data augmentation method as a cropping approach. The training images are firmly cropped with a patch size of 130 × 130 and size of crop is 130.

From the LUNA 16 database, we obtained 900 CT images using 20 types of different scanners from about 1100 lung cancer cases. Similar to the above pixel range, the obtained CT images also have a 512 × 512 pixel range with 2.5 mm slice resolution. The training and testing ratio for CT images is 80:20 (i.e., 720 images for training and 180 images for testing) in which 80% is utilised for training and 20% for testing. For further improving the training image set, we utilise cropping strategy as data augmentation as similar to the above method. We crop the image patch to 130 × 130 and crop size is 130.

The acquired images are fed into the designed model, and we perform normalisation using min–max normalisation method [41]. The formulation of image normalisation is provided as below:

$$\mathrm{Im}_{\mathrm{Nor}} = (\mathrm{Img} - \min)\frac{\mathrm{newMax} - \mathrm{newMin}}{\mathrm{Max} - \mathrm{Min}} + \mathrm{newMin}. \quad (1)$$

From the above equation, $\mathrm{Im}_{\mathrm{Nor}}$ denotes the normalised image, Max and Min denotes the maximum and minimum intensity values of the original image, whereas the newMax and newMin denotes the intensity values of the normalised image. The normalised image is then fed to the designed model for the segmentation task.

4 | Proposed Methodology

Our proposed research adopts three imaging modalities named CT, MRI and SPECT, respectively, with two major modules. Module (a) is named as the versatile scale malleable image integration-patch wise attention network and module (b) backbone (lite wave transformer). Module (a) accepts input as CT and MRI images which are then undergone patch cropping, patch analysis, patch fusion and attention map generation. Module (b) accepts input as attention map and SPECT image to the backbone module which undergoes robust image analysis for image classification into two classes normal and malignant. Our designed model adopts the spatial information of the images in patch wise manner to conclude the detection with better accuracy. Figure 1 represents the pipeline of the proposed model. The decision for adopting PANet and LWT is from their unique strengths in handling the multi modal images. PANet was chosen because its patch-wise approach allows for localised attention on small image regions, enabling the model to focus on fine details which is crucial for detecting tumours with irregular boundaries. LWT, on the other hand, excels at capturing spectral variations and functional information from SPECT images which is vital for differentiating between active and non-active tumour regions. These methods were selected over others because they provide a balance between local detail extraction and global feature integration, ensuring precise segmentation in multi-modal contexts such as CT, MRI and SPECT. The key innovation of every module is shown below:

- *Scale-malleable convolution layer (SMCL):* Unlike traditional convolution layers, SMCL is designed to adapt to different scales in medical imaging. It may dynamically change the receptive field to retain both local and global context, capturing minute details in tumour borders as well as larger anatomical features. This adaptability offers SMCL a particular advantage for dealing with the complicated scale changes encountered in multi-modal medical imaging such as CT, MRI and SPECT.
- *Patch-wise attention network (PANet):* PANet is a revolutionary method of applying attention that focuses on specific portions of a picture. Unlike classic attention techniques, which tend to focus on bigger or global picture areas, PANet improves the model's capacity to collect fine-grained features inside each patch. This concentrated focus enables PANet to better distinguish between tumour borders and adjacent tissues, resulting in higher segmentation precision than typical attention techniques that work at a larger level.
- *Light wave transformer (LWT):* LWT works with PANet to create an efficient and lightweight attention mechanism designed exclusively for multi-modal picture integration. Whereas standard transformers can be computationally expensive, LWT is designed to handle the unique peculiarities of medical imaging data, resulting in effective feature extraction without a significant computational burden. LWT's capacity to combine spectral and spatial information from CT, MRI and SPECT images distinguishes it from other attention processes, enhancing accuracy and computing efficiency.

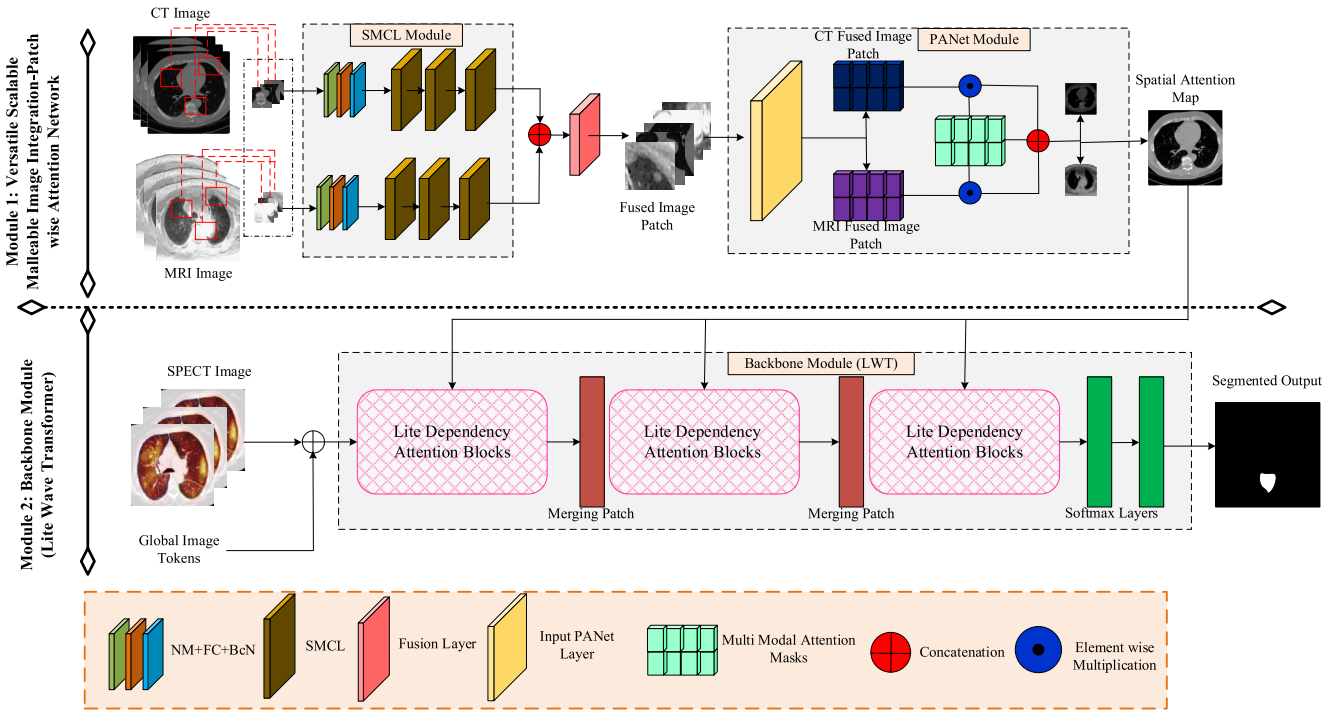


FIGURE 1 | Overall architecture of the proposed work.

4.1 | Versatile Scale Malleable Image Integration-Patch Wise Attention Module

Under this module, there are three sub-modules involved for the image analysis such as (i) image cropping module, (ii) scaled malleable convolution layers (SMCL) module and (iii) patch wise attention network (PANet) module. The detailed explanation of those modules is explained below:

4.1.1 | Image Cropping Module

The two input images CT $(\mathrm{Img}_{\mathrm{CT}})$ and MRI $(\mathrm{Img}_{\mathrm{MRI}})$ are given to the parallel image cropping module which iterates all the portions of the images and crops the image patches of equal sizes (10×10 pixel sizes). The formulation of cropping processes in the image cropping module is provided below:

$$\mathrm{ICM}(\mathrm{Img}_{\mathrm{CT}}) = \mathrm{Pc}^1(\mathrm{Img}_{\mathrm{CT}}), \mathrm{Pc}^2(\mathrm{Img}_{\mathrm{CT}}), \ldots, \mathrm{Pc}^N(\mathrm{Img}_{\mathrm{CT}}), \quad (2)$$

$$\mathrm{ICM}(\mathrm{Img}_{\mathrm{MRI}}) = \mathrm{Pc}^1(\mathrm{Img}_{\mathrm{MRI}}), \mathrm{Pc}^2(\mathrm{Img}_{\mathrm{MRI}}), \ldots, \mathrm{Pc}^N(\mathrm{Img}_{\mathrm{MRI}}). \quad (3)$$

From the above Equations (2) and (3), $\mathrm{ICM}(\mathrm{Img}_{\mathrm{CT}}, \mathrm{Img}_{\mathrm{MRI}})$ denotes the input of image cropping module as CT and MRI and $\mathrm{Pc}^1(.)$ denotes the equal size image patches from CT and MRI images, respectively.

4.1.2 | Scaled Malleable Convolution Layer (SMCL)

The cropped patches of the CT and MRI images are fed to the SMCL, and the patches are then processed to normal module (NM) in flexible convolution (FC) layer, batch normalisation (BcN) layer, and ReLU layer. The formulation of those layers can be defined as follows:

$$\mathrm{Pc}_{\mathrm{NM}}^{\mathrm{Op}} = \mathrm{ReLU}\left(\mathrm{BcN}\left(\mathrm{FC}\left(\mathrm{Pc}_{\mathrm{NM}}^{\mathrm{ip}}\right)\right)\right). \quad (4)$$

From the above equation, $\mathrm{Pc}_{\mathrm{NM}}^{\mathrm{Op}}$, and $\mathrm{Pc}_{\mathrm{NM}}^{\mathrm{ip}}$ denotes the output and input of image patches from the normal module. It is to be noted that the patches are obtained and processed in parallel to continually obtain the functional, structural and metabolic details from both the CT and MRI images patches, respectively [42]. Therefore, the information perseveration is necessary for enabling better tumour segmentation. For better preserving the context information in an image patch, we adopt parallel SMCLs which replace the vanilla convolutional layers with flexible processes for image patches and obtains the context information. There are three SMCL involved parallelly for CT and MRI images patches, respectively. Other than that, the adoption of SMCL is to firmly extract the deep and latent features, respectively, with the deep analysis capability. The operations involved in the SMCLs are provided below:

$$\mathrm{Pc}_{\mathrm{SMCL1}}^{\mathrm{Op}} = \mathrm{SLA}\left(\mathrm{Nor}^{\mathrm{La}}\left(\mathrm{Pc}_{\mathrm{SMCL}}^{\mathrm{ip}}\right)\right) + \mathrm{Pc}_{\mathrm{SMCL}}^{\mathrm{ip}}. \quad (5)$$

From the above equation, $\mathrm{Pc}_{\mathrm{SMCL1}}^{\mathrm{Op}}$ denotes the initial additive operation of the SMCL, $\mathrm{Pc}_{\mathrm{SMCL}}^{\mathrm{ip}}$ denotes the input to the SMCL, $\mathrm{Nor}^{\mathrm{La}}$ denotes the normalisation layer and SLA denotes the self-attention layer. The final additive operation of SMCL can be formulated as follows:

$$\mathrm{Pc}_{\mathrm{SMCL}}^{\mathrm{Op}} = \mathrm{ANN}\left(\mathrm{Nor}^{\mathrm{La}}\left(\mathrm{Pc}_{\mathrm{SMCL1}}^{\mathrm{Op}}\right)\right) + \mathrm{Pc}_{\mathrm{SMCL1}}^{\mathrm{Op}}. \quad (6)$$

From the above equation, the output of the SMCLs can be denoted as Pc^{Op}_{SMCL}, whereas the ANN denotes the artificial neural network. Finally, the output from the two parallel SMCLs are provided to the FC and tanh function to obtain the fused image patches as $Pc^{Op}_{SMCL}[CT] \oplus Pc^{Op}_{SMCL}[MRI] = Fu_{Pc}(CT, MRI)$.

Since we design a multi-modal fused patch output from the CT and MRI images which provides both the functional and operational information in a patch wise manner, the loss function associated with them can also be devised to enhance the training rate. The total loss function can be formulated as follows:

$$Tot^{loss} = Fun^{loss} + Ope^{loss}. \tag{7}$$

From the above equation, the Fun^{loss} and Ope^{loss} denotes the functional and operational losses, respectively. The consideration of Ope^{loss} is to ensure the amount of operational information in the fused patch which can also be computed using operational similarity measure (OSM). The OSM provides the similarity among the source and fused image patch, and it can be formulated as follows:

$$\begin{aligned} Ope^{loss} = 1 - OSM\left(Fu_{Pc}, Pc^{Op}_{SMCL}[CT]\right) \\ + \beta\left(1 - OSM\left(Fu_{Pc}, Pc^{Op}_{SMCL}[MRI]\right)\right). \end{aligned} \tag{8}$$

From the above equation, the balancing factor among the Fu_{Pc} and source image patch ($\forall$), that is, CT or MRI is denoted as β. The formulation of OSM can be computed as follows:

$$OSM(\forall, Fu_{Pc}) = \frac{\left(2_{\partial_\forall \partial_{Fu_{Pc}}} + \beth_1\right)\left(2_{\rho_{\forall Fu_{Pc}}} + \beth_2\right)}{\left(\partial^2_\forall + \partial^2_{Fu_{Pc}} + \beth_1\right)\left(\rho^2_\forall + \rho^2_{Fu_{Pc}} + \beth_2\right)}. \tag{9}$$

From the above equation, $\partial_\forall$ and $\partial_{Fu_{Pc}}$ denotes the mean of source image patch and fused image patch. The $\rho^2_\forall$ and $\rho^2_{Fu_{Pc}}$ denotes the variance of source and fused image patch, respectively, and the problem of instability can be resolved by adopting constants $\beth_1$ and $\beth_2$.

In order to further resolve the problem of transfer artefacts (i.e., from the source image patch, the region information can be transferred to the fused image patch), we adopt restriction measures named mutual information restriction measure (MIRM) in the structural level and can be formulated as follows:

$$\begin{aligned} Fun^{loss} = \varphi MIRM\left(Fu_{Pc}, Pc^{Op}_{SMCL}[CT]\right) \\ + \aleph MIRM\left(Fu_{Pc}, Pc^{Op}_{SMCL}[MRI]\right). \end{aligned} \tag{10}$$

From the above equation, the controlling parameters can be denoted as φ and $\aleph$. The formulation of MIRM can be denoted as follows:

$$\begin{aligned} MIRM(\forall, Fu_{Pc}) = \propto CEL(\forall, Fu_{Pc}) \\ + (1 - \propto)\frac{1}{Bz}\sum_{bz=1}^{Bz}\left(-Pc^{bz}_{lo}(\forall, Fu_{Pc})\right). \end{aligned} \tag{11}$$

From the above equation, the CEL denotes the cross-entropy loss among $\forall$ and Fu_{Pc}, the $\propto$ denotes the weight value and ranges from [0, 1]. The size of the batch can be denoted as Bz, and the functional information lower bound can be denoted as $-Pc^{bz}_{lo}$.

4.1.3 | Patch Wise Attention Network (PANet)

Once the $Fu_{Pc}(CT, MRI)$ is obtained, it is then provided to the PANet module, and its dimensions are represented as $Fu_{pa(CT|MRI)}[WHDC]$ width, height, depth and channel, respectively [43]. The $Fu_{pa(CT|MRI)}[WHDC]$ is then subjected to the averaging pooling for squeezing the dimensions to get the 1D-vector that can be represented as $(\Lambda_{CT|MRI}(2\mathbb{C}))$. For obtaining the channel weights for multi-modal fused patches $Fu_{pa(CT|MRI)}$, the following equation can be computed as follows:

$$\mathbb{S}_{CT|MRI}(2\mathbb{C}) = \sigma\left(wei^2 \cdot \tau\left(wei^1 \cdot \Lambda_{CT|MRI}(2\mathbb{C})\right)\right), \tag{12}$$

where τ and σ determines the ReLu and sigmoid function for the $\mathbb{S}_{CT|MRI}(2\mathbb{C})$ and wei^1 and wei^2 denotes the weight parameters matrices. After that, we accomplish multiplication in element wise manner to get the $Fu'_{pa(CT|MRI)}[WHDC]$ and can be formulated as follows:

$$\begin{aligned} Fu'_{pa(CT|MRI)}[WHDC] = \mathbb{S}'_{CT|MRI}[WHDC] \\ \times Fu_{pa(CT|MRI)}[WHDC]. \end{aligned} \tag{13}$$

In addition to that, the obtained result $Fu'_{pa(CT|MRI)}[WHDC]$ is then sliced towards channel axis to produce the separate multi-modal channel weights for both the CT and MRI, respectively. The formulation is provided as below:

$$Fu'_{CT}[WHDC] = Fu'_{pa(CT|MRI)}[WHDC], \tag{14}$$

$$Fu'_{MRI}[WHDC] = Fu'_{pa(CT|MRI)}[WHDC]. \tag{15}$$

From the above equation, Fu'_{CT} and Fu'_{MRI} determines the separate multi modal channel weights for CT and MRI.

$Fu'_{pa(CT|MRI)}[WHDC]$ is then provided to the 4D convolutional kernels with the sigmoid function to get the masked spatial attention maps for CT and MRI respectively, and can be formulated as follows:

$$\mathbb{T}_{CT}(WHD) = \sigma\left(4D\ kernel_{CT} \times Fu'_{CT+|MRI+}(WHDC)\right), \tag{16}$$

$$\mathbb{T}_{MRI}(WHD) = \sigma\left(4D\ kernel_{CT} \times Fu'_{CT+|MRI+}(WHDC)\right). \tag{17}$$

The above Equations (16) and (17) are obtained by performing the 4D convolutional and squeezing operations respectively in which the $\times$ is denoted as the convolutional symbol. Finally, the feature insisted attention map can be obtained by performing element wise and can be formulated as three Equations (18–20) below:

$$Fu''_{CT}(WHDC) = \mathbb{T}_{CT}(WHDC) \times Fu'_{CT}(WHDC), \tag{18}$$

$$Fu''_{MRI}(WHDC) = \mathbb{T}_{MRI}(WHDC) \times Fu'_{MRI}(WHDC), \quad (19)$$

where the $Fu''_{CT}(WHDC)$ and $Fu''_{MRI}(WHDC)$ are the separate polished feature fusion maps from the $Fu'_{CT}[WHDC]$ and $Fu'_{MRI}[WHDC]$, respectively. The final feature maps can be obtained by fusing the refined CT and MRI feature maps and can be formulated as follows:

$$Fea_{Fu}(WHDC) = Fu''_{CT}(WHDC) + Fu'_{MRI}[WHDC]. \quad (20)$$

The above Equation (20) shows the feature insisted attention maps obtained from the CT and MRI image patches.

4.2 | Lite Wave Transformer (LWT)

The feature insisted attention maps obtained from the previous module are then provided to the backbone module named LWT which composed of three consecutive lite dependency attention blocks (LDAB) for ensuring self-local attention, global dependency broadcast and global accumulation as shown in Figure 2. To be clearer, the LWT effectively captures the variation in SPECT by focusing the functional heterogeneity, metabolic activities and suppressing the background noises. This allows the LWT to robustly identify the tumour regions which might not presented in the CT or MRI such as structural images. Furthermore, the SPECT provides functional boundaries based on the structural data complementation, metabolic changes and guarantees that the segmentation is both functionally and anatomically accurate. The SPECT integration in LWT leads to delineation in tumours particularly in complex heterogenous tumour activities. $Fea_{Fu}(WHDC)$ is provided as an input to the LWT. The input $Fea_{Fu}(WHDC)$ is forwarded to the LDAB. We also provide global image patch tokens as an input to the LDAB [44].

In LDAB, there are three parallel attention layers involved, such as self-local attention, global dependency broadcast and global accumulation, respectively. The detailed mathematical explanation of those attentions involved are provided below:

4.2.1 | Self-Local Attention

$Fea_{Fu}(WHDC)$ is provided as an input to the self-local attention layer. From $Fea_{Fu}(WHDC)$, we split the input into non-concurrent windows of shape $\left(\frac{He}{Ws} \times \frac{Wi}{Ws}, Ws \times Ws, C\right)$ in which He, Wi, Ws and C denote the height, width, window size and channels, respectively. For every Ws of the input, we apply self-local attention and can be formulated as follows:

$$\begin{aligned} Fea_{Fu}{}^{Self-local} &= Att\left(Fea_{Fu}{}^{q}, Fea_{Fu}{}^{k}, Fea_{Fu}{}^{v}\right) \\ &= Smax\left(Fea_{Fu}{}^{q} Fea_{Fu}^{k\ T}\right) Fea_{Fu}{}^{v}. \end{aligned} \quad (21)$$

From the above equation, the projections of Q, K and V provide the $Fea_{Fu}{}^{q}, Fea_{Fu}{}^{k}, Fea_{Fu}{}^{v}$. Henceforth, we obtain lesser

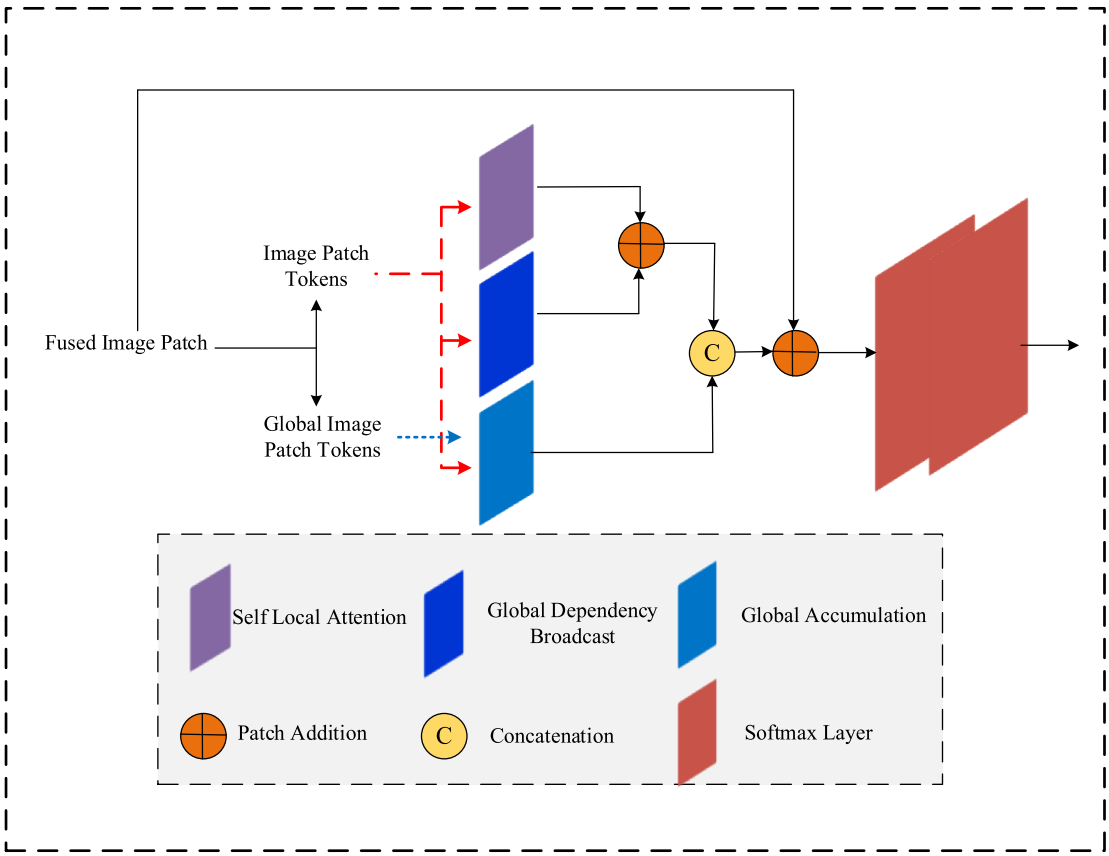


FIGURE 2 | Lite dependency attention blocks in LWT.

computation complexity of $\left(\frac{He}{Ws} \times \frac{Wi}{Ws}\right) \times (Ws \times Ws)^2 = He \times Wi \times Ws \times Ws$.

4.2.2 | Global Accumulation Layer

In order to obtain global information from the image feature maps, the proposed work designs a global embedding mode which is denoted as $\mathfrak{G} \in \mathbb{R}^{Tr \times C}$ in which Tr denotes the transpose operation. The designed $\mathfrak{G}$ is utilised for performing broadcasting and accumulating the global information. Initially, the designed $\mathfrak{G}$ concatenates the global representation image information from the input feature map. After that, it broadcasts that information to the feature maps in a homogenous manner. In addition to that, for computing the global accumulation output, we utilise both the self-local attention and accumulation input using $\mathfrak{G}$ and Fea_{Fu} which can be formulated as follows:

$$\widehat{GAc} = Att(\mathfrak{G}_q, Fea_{Fu}{}^k, Fea_{Fu}{}^v). \tag{22}$$

From the above equation, $\widehat{GAc}$ is denoted as the brand new token and can be utilised for further processes.

4.2.3 | Global Dependency Broadcast

In order to enhance the global dependencies in the image features, we adopt global information from the image token which are then broadcasted by utilising $\widehat{GAc}$ and can be formulated as follows:

$$Fea_{Fu}{}^{Global} = Att\left(Fea_{Fu}{}^q, \widehat{GAc}^k, \widehat{GAc}^v\right). \tag{23}$$

The image patch tokens obtained are calculated in an element-wise manner and can be formulated as follows:

$$Fea_{Fu}{}^{New} = Fea_{Fu}{}^{local} + Fea_{Fu}{}^{Global}. \tag{24}$$

The above equation aggregates the local and global information from the feature maps and provides the output in terms of three classes such as normal, benign and malignant.

5 | Experimental Results

This section comprehensively presents the outcomes of our innovative approach, both in terms of quantitative and qualitative assessments when compared to existing methodologies. Additionally, it provides a thorough evaluation of our approach's performance across the distinct datasets in comparison to existing methods. The section begins by offering an insight into the characteristics of the datasets, followed by a detailed examination of results across different scenarios, and culminates in a meaningful and insightful discussion of the findings.

5.1 | Performance Assessment

The lung cancer segmentation and detection model named as VSMI^2 – PANet (versatile scale malleable image integration-patch wise attention network) was executed using MATLAB on an HP Pavilion laptop featuring an AMD Ryzen 5 5600H processor with Radeon Graphics, operating at 3.30 GHz. The computer boasts a random-access memory (RAM) capacity of approximately 8 GB and runs on the Windows 11 operating system (OS). To assess its performance, we conducted a thorough validation by comparing it with established methods, utilising two well-regarded benchmark datasets, named Harvard and LUNA 16. The existing methodologies used for benchmarking and performance evaluation are detailed below:

- MSANet [22]: This work adopts PET and CT images for segmenting the lung tumour using deep learning entrenched spatial attention mechanism.
- MM3D-Lung [23]: In this work, both CT and PET images were utilised for detecting the lung tumour using 3D attention mechanism.
- Cross-MM [24]: Performing detection of lung tumours from multi modal MRI images in cross domain manner.
- EMM-LC [26]: Fusing both the clinical data and medical CT image for segmenting the lung tumour using multi modal entrenched deep learning algorithm.
- DL-MGAM [27]: Predicting the severity lung cancer tumours by exploiting multi-modal attention-based model using deep learning algorithm.
- Conv-UNet [31]: Adopting computed tomography image for detecting the lung tumour using combination of dilated convolution and UNet frameworks.

The performance metrics utilised for validating the proposed VSMI^2 – PANet and existing works are accuracy, precision, recall, F-measure and AUC curve.

Table 1 shows how the proposed VSMI^2-PANet model performs across several imaging modalities. CT-based segmentation has the highest accuracy (95.66%) because of its thorough anatomical imaging but a slightly lower precision (94.00%) than MRI and SPECT. MRI has a more balanced performance with a little lower accuracy (95.35%) but a better *F*1-Score (92.00%) because of improved soft tissue contrast. SPECT has the highest precision (94.65%), but slightly lower accuracy (95.19%) because of its emphasis on functional imaging, which supplements the anatomical features of CT and MRI. The model without LWT performs well overall, but with a slightly lower accuracy (94.16%) and *F*1-Score (92.02%), demonstrating LWT's increased value for feature extraction. VSMI^2-PANet with LWT improves segmentation accuracy by capturing spectral and spatial fluctuations, especially when combining complementary data from CT, MRI and SPECT (94.65% precision, 93.53% recall).

Additionally, we conduct a comparative analysis between our proposed approach and a selection of traditional machine learning and deep learning classifiers including convolutional neural networks (CNN), K-nearest neighbours (KNN), artificial

TABLE 1 | Comparison within proposed models.

Proposed model	Performance metrics				
	Accuracy (%)	Precision (%)	Recall (%)	*F*1-score (%)	AUC (%)
VSMI^2 – PANet with CT	95.66	94.00	93.23	91.76	92.34
VSMI^2 – PANet with MRI	95.35	94.23	93.00	92.00	92.88
VSMI^2 – PANet with SPECT	95.19	94.65	93.53	91.23	92.26
VSMI^2 – PANet without LWT	95.27	94.16	93.09	92.02	92.28
VSMI^2 – PANet with LWT	95.19	94.65	93.53	91.23	92.26

neural network (ANN) and support vector machine (SVM). The performance metrics used for evaluation are detailed in the following section.

$$\text{Accuracy (Acc)} = \frac{\text{TrP} + \text{TrN}}{\text{TrP} + \text{TrN} + \text{FlP} + \text{FlN}} \times 100, \quad (25)$$

$$\text{Precision (Pre)} = \frac{\text{TrP}}{\text{TrP} + \text{FlP}} \times 100, \quad (26)$$

$$\text{Recall (Rec)} = \frac{\text{TrP}}{\text{TrP} + \text{FlN}} \times 100, \quad (27)$$

$$F\text{-Measure (FM)} = 2\frac{\text{Pre} \times \text{Rec}}{\text{Pre} + \text{Rec}} \times 100. \quad (28)$$

From the above Equations (25–28), TrP, TrN, FlP and FlN denote the true positive, true negative, false positive and false negative rates, respectively. The AUC-ROC curve offers a comprehensive assessment of the classifier's classification performance across a range of possible thresholds. A higher AUC signifies superior performance by the classifier. Table 1 shows the comparison of proposed model in three different image modalities, and also, we show the performance of the model with and without backbone interpretation.

5.2 | Evaluation of LUNA 16 Database

For assessing the performance of the proposed model performance on LUNA 16 database, we perform the comparison of both the conventional ML/DL models and existing designed models for lung tumour from CT images. It is to be noted that we adopt both the augmented and non-augmented data for examining the performance of the proposed model.

Tables 2–4 and Figure 3 show the comparison of the performance of proposed and conventional ML/DL models on the CT image. The conventional ML/DL models utilised for performance assessment are SVM, KNN, CNN and ANN. In addition to that, we have also examined the sensitivity separately for the proposed and conventional models in five pre-defined false positive rates: 1/8, 1/4, 1/2, 2 and 4. From the pre-defined false positive rates, we additionally determine the viable performance metrics (VPM) and free-response receiver operating curve (FROC) to additionally examine the performance of the proposed model.

In a similar manner, the proposed model is also compared with the state of the works such as [22, 23, 26, 31]. As aforesaid, we have utilised general performance metrics such as accuracy, precision, recall, *F*1-score and AUC curve for evaluation. Furthermore, we have also utilised five pre-defined values such as 1/8, 1/4, 1/2, 2 and 4

From the analysis of graphical and tabulation results, the proposed work achieves better results than state-of-the-art works. In addition to that, we also show the comparison of the FPR with the different pre-defined sensitivity values. The reason for such better performance of the proposed model is because we adopt parallel layer mechanisms for processing the CT images in patch wise manner. By performing the parallel layer processing in patch wise manner, the problem of contrast and ionising problems gets reduced which in turns gains higher accuracy by using CT image.

5.3 | Evaluation of Harvard Database

From the Harvard database, we acquire both the MRI and SPECT images for examining the severity of the lung tumour. Similar to the previous comparative analysis models, this section also explains the performance of the proposed model with the MRI and SPECT images, respectively.

5.3.1 | Analysis Using MRI Images

The utilised database provides the clear picture of lung tumours along with the ground truth labels. Once performing preprocessing on MRI images in terms of noise removal and normalisation, the differences and characteristics variability in the MRI images enables higher variance in the designed model performance.

Tables 5–7 and Figure 4 show the comparison of proposed model performance with the conventional ML/DL models, such as SVM, KNN, CNN and ANN. Furthermore, we have analysed the performance of the proposed model with the VPM and FROC respectively with five different FPR rates listed as 1/8, 1/4, 1/2, 2 and 4. Furthermore, the table shows the comparison of proposed model performance with the existing works are [24, 25, 27, 45]. In similar manner, the performance of the proposed and existing models is assessed by adopting performance metrics measures

TABLE 2 | Comparison of proposed versus conventional models with CT images.

Models	Accuracy (%)	Precision (%)	Recall (%)	*F*1-score (%)	AUC
SVM	85.97	86.00	85.85	85.78	0.8627
KNN	86.17	87.76	87.29	87.08	0.8717
CNN	90.19	92.98	91.75	91.68	0.9017
ANN	92.99	93.56	93.98	94.86	0.9378
Proposed model	98.79	98.83	98.93	99.18	0.9837

TABLE 3 | Comparison of proposed versus existing works with CT images.

Models	Accuracy (%)	Precision (%)	Recall (%)	*F*1-score (%)	AUC
[22]	86.89	86.90	87.23	87.26	0.8726
[23]	87.77	87.56	87.39	87.78	0.8828
[26]	92.20	93.98	93.95	93.76	0.9327
[31]	94.86	94.26	94.26	94.78	0.9580
Proposed model	97.09	99.83	99.23	99.26	0.9915

TABLE 4 | FPR and VPM analysis for proposed versus existing models.

	Sensitivity rate with false positive rates on conventional models					
Models	**1/8 (%)**	**1/4 (%)**	**1/2 (%)**	**2 (%)**	**4 (%)**	**VPM (%)**
SVM	93.17	92.03	92.05	92.29	92.63	40.79
KNN	94.06	94.36	94.79	94.92	94.00	47.26
CNN	96.78	96.68	95.85	96.89	97.45	58.00
ANN	97.01	96.99	97.36	96.01	97.07	79.89
Proposed model	99.13	99.97	98.89	98.03	99.85	89.11
	Sensitivity rates with false positive rates on existing works					
Methods	**1/8 (%)**	**1/4 (%)**	**1/2 (%)**	**2 (%)**	**4 (%)**	**VPM (%)**
[22]	94.77	94.03	93.05	94.29	93.63	50.08
[23]	95.06	95.36	95.79	95.92	95.56	58.45
[26]	97.78	96.47	96.98	96.03	96.27	65.00
[31]	98.00	98.11	98.47	96.12	96.18	87.37
Proposed model	99.24	99.08	98.90	98.14	99.96	92.45

such as accuracy, precision, recall, *F*1-score and AUC curve and pre-defined sensitivity values 1/8, 1/4, 1/2, 2 and 4.

By examining both graphical representations and tabulated data, it is evident that our proposed approach outperforms current state-of-the-art methods. To clarify further, the table and figures clearly illustrate that the proposed method exhibits superior performance compared to traditional ML/DL models. The detailed breakdown of performance metrics for both our proposed method and existing models/techniques highlights the significant advantage of our approach. Additionally, we present a comparison of false positive rates (FPR) across various pre-defined sensitivity values. The improved performance of our model can be attributed to our innovative use of parallel layer mechanisms for processing MRI images in a patch-wise manner. This approach effectively mitigates time consumption and complexity issues, resulting in a higher accuracy when working with MRI images.

5.3.2 | Analysis Using SPECT Images

The adopted images are pre-processed and refined to improve the effectiveness of the designed model in terms of variability and characteristics in the SPECT images. The Tables 8–10 and Figure 5 present a comprehensive evaluation of our proposed model's performance when compared to conventional ML/DL models, including SVM, KNN, CNN and ANN. For instance, SVM is known for its versatility in handling various data types while KNN relies on proximity-based classification. CNNs excel in image-related tasks due to their convolutional layers while ANNs offer a broad range of applications. By comparing the proposed model's performance to these models, we gain valuable insights into its strengths and weaknesses.

Additionally, we conducted a thorough analysis of our model's performance using the visual pattern matching (VPM) and

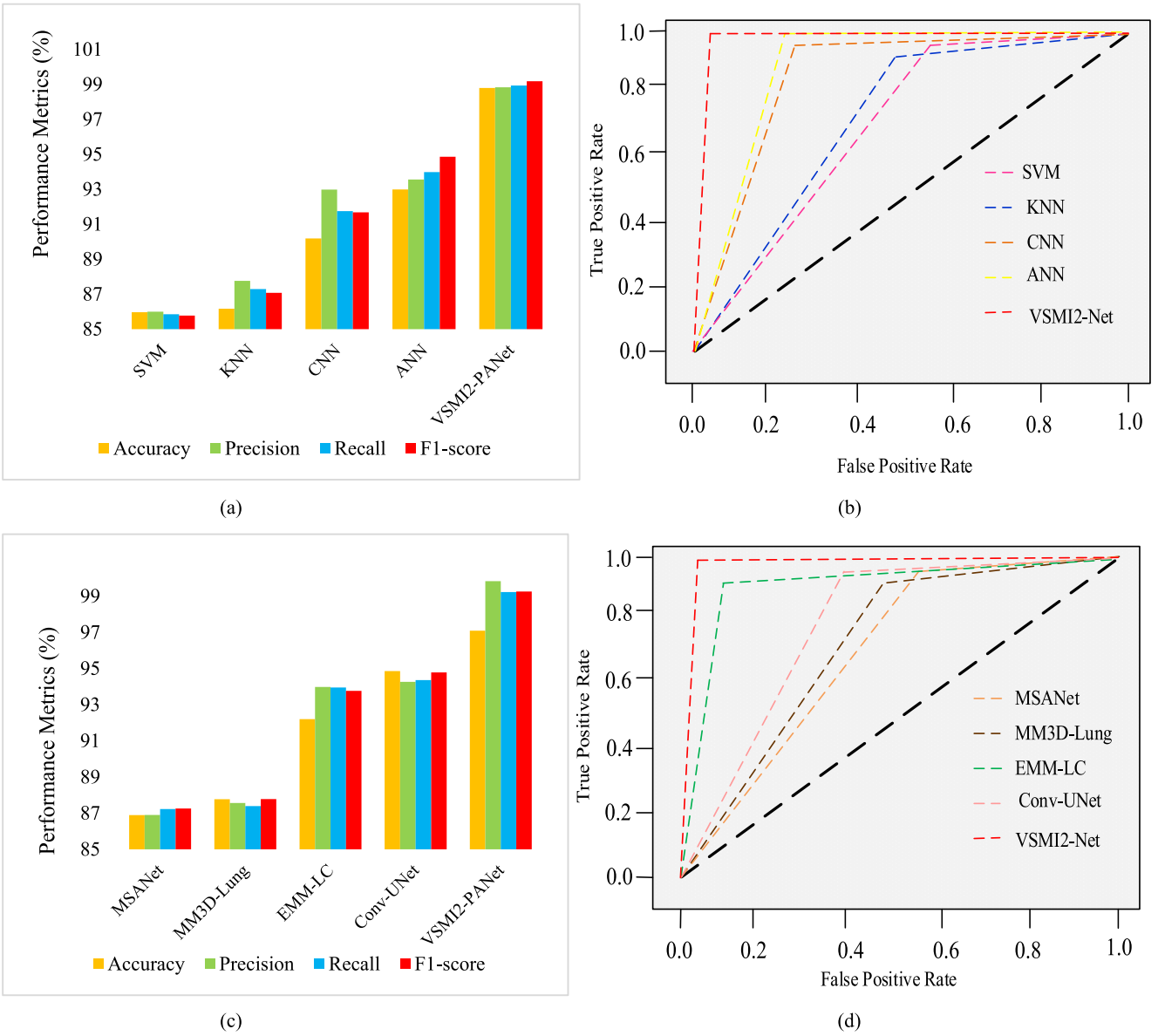


FIGURE 3 | (a) Comparison of proposed versus conventional models using CT images. (b) Comparison of proposed versus existing works using CT images. (c) AUC comparison on proposed versus existing models using CT images. (d) AUC comparison of proposed versus existing models using CT images.

TABLE 5 | Comparison of proposed versus conventional models with MRI images.

Models	Accuracy (%)	Precision (%)	Recall (%)	*F*1-score (%)	AUC
SVM	85.00	85.15	85.45	85.70	0.8580
KNN	86.17	86.52	86.74	86.95	0.8695
CNN	88.20	86.50	86.80	87.00	0.8817
ANN	90.99	91.99	91.79	91.59	0.9193
Proposed model	98.00	98.09	99.03	99.06	0.9937

free-response receiver operating characteristic (FROC) methods, considering five distinct false positive rate (FPR) levels: 1/8, 1/4, 1/2, 2 and 4. FROC, on the other hand, measures the model's ability to detect lesions or anomalies across various sensitivity levels. These analyses provide a nuanced understanding of how the proposed model performs under different conditions, shedding light on its practical applicability in real-world scenarios. Furthermore, the same table and

TABLE 6 | Comparison of proposed versus existing works with MRI images.

Models	Accuracy (%)	Precision (%)	Recall (%)	*F*1-score (%)	AUC
[24]	86.90	86.01	87.34	87.37	0.8837
[27]	88.66	86.45	87.28	87.78	0.8738
[25]	93.32	95.20	95.23	90.54	0.9579
[45]	96.08	96.48	96.48	96.90	0.9702
Proposed model	99.22	99.05	99.45	99.48	0.9937

TABLE 7 | Comparison of FRP and VPM with MRI images.

	Sensitivity rate with false positive rates on conventional models					
Models	**1/8 (%)**	**1/4 (%)**	**1/2 (%)**	**2 (%)**	**4 (%)**	**VPM (%)**
SVM	93.97	96.13	96.13	93.99	94.89	49.92
KNN	96.06	95.36	95.94	97.07	93.00	54.58
CNN	93.78	94.68	92.85	92.89	95.45	60.44
ANN	95.01	92.99	95.36	95.01	97.07	85.92
Proposed model	98.13	98.97	99.89	99.03	98.85	92.74
	Sensitivity rates with false positive rates on existing works					
Methods	**1/8 (%)**	**1/4 (%)**	**1/2 (%)**	**2 (%)**	**4 (%)**	**VPM (%)**
[24]	94.17	95.23	95.46	95.75	95.07	60.20
[27]	95.37	96.59	96.80	96.17	96.00	70.40
[25]	93.42	94.89	97.39	96.89	97.45	80.60
[45]	96.07	96.19	98.36	97.09	98.07	90.80
Proposed model	99.13	99.06	99.22	99.44	99.85	95.00

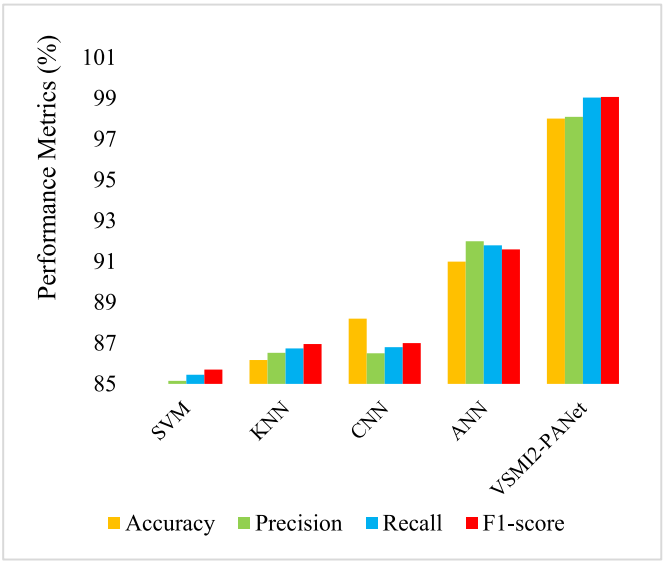
figures display a performance comparison between our proposed model, specifically referencing sources [46–48]. Similarly, we assessed the performance of both our proposed model and the existing ones by employing a range of performance metrics, including accuracy, precision, recall, *F*1-score and the area under the curve (AUC), while considering pre-defined sensitivity values of 1/8, 1/4, 1/2, 2 and 4.

Upon a comprehensive analysis of both visual representations and structured data, it becomes apparent that our suggested strategy surpasses the capabilities of current state-of-the-art techniques. To elaborate further, the data presented in the tables and visual aids unequivocally demonstrates that our proposed methodology exhibits markedly superior performance when compared to the conventional machine learning and deep learning models. These visual aids offer an exhaustive breakdown of performance metrics for both our innovative approach and the established models and methods, underscoring the significant edge of our approach. Additionally, we offer a comparative assessment of false positive rates (FPR) across a spectrum of predetermined sensitivity thresholds. The enhanced performance of our model can be attributed to our inventive implementation of parallel layer mechanisms tailored for the patch-wise processing of SPECT images in both the feature analysis and backbone model, respectively. This approach effectively mitigates concerns pertaining to time efficiency and intricacy, ultimately culminating in elevated precision when dealing with SPECT images.

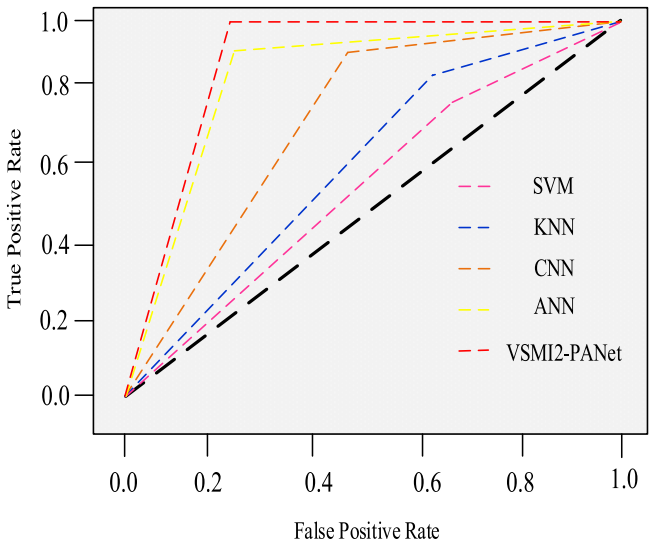
6 | Ablation Studies

The proposed model is further evaluated at every stage by exploiting FROC. From Figure 6, we try to illustrate the presentation of the proposed research with and without LWT, respectively. At first stage, the effectiveness of the proposed model is compared without LWT in terms of number of FPR scans. From the graphical results, it is seen that for the less average number of FPR per scan, the rate of sensitivity for the proposed model without LWT is lesser. Conversely, when the proposed work is combined with LWT, the rate of sensitivity gets higher although the number of FPR scans increases to its maximum limit.

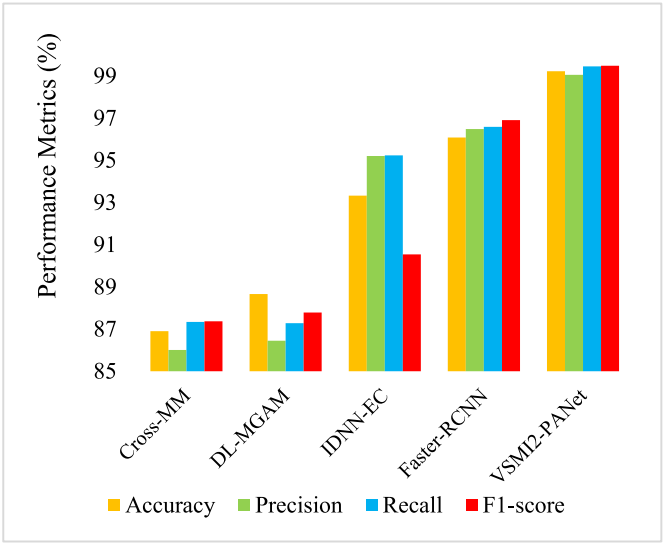
To be clearer, when the number of FPR is at initial stage (i.e., 0.15), the sensitivity achieved by the proposed work without LWT is only 0.65%, whereas the proposed work with LWT achieves 0.85%. Additionally, by computing the VPM scores for the both the scenarios at various FPR pre-defined scenarios, the value of VPM performed better in the proposed work with LWT and performed worse in the proposed work without



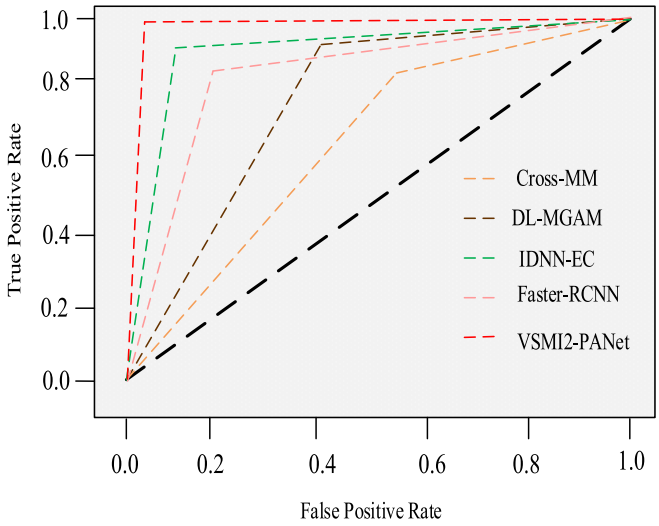
(a)



(b)



(c)



(d)

FIGURE 4 | (a) Comparison of proposed versus conventional models using MRI images. (b) Comparison of proposed versus existing works using MRI images. (c) AUC comparison on proposed versus existing models using MRI images. (d) AUC comparison of proposed versus existing models using MRI images.

TABLE 8 | Comparison of proposed versus conventional models with SPECT images.

Models	Accuracy (%)	Precision (%)	Recall (%)	*F*1-score (%)	AUC
SVM	87.22	87.27	87.67	87.92	0.8792
KNN	88.39	88.74	88.96	88.27	0.8827
CNN	90.42	88.70	90.02	90.22	0.9039
ANN	92.22	93.22	93.92	93.72	0.9325
Proposed model	99.33	99.34	99.25	99.28	0.9959

LWT. The reason for such poor performance over without LWT is because the stringent analysis of spatial information by utilising local and global attention is lacked in without LWT scenario.

To show the effectiveness of the proposed work in a clear picture, Figures 7–9 show the comparison of proposed model with and without LWT utilised in three imaging modalities named CT, MRI and SPECT images, respectively. Since the SPECT

TABLE 9 | Comparison of proposed versus existing works with SPECT images.

Models	Accuracy (%)	Precision (%)	Recall (%)	*F*1-score (%)	AUC
[46]	86.57	86.56	86.40	87.90	0.8849
[47]	92.37	93.49	93.99	92.54	0.9579
[48]	95.20	95.99	97.45	97.03	0.9611
Proposed model	99.92	99.95	99.97	99.49	0.9999

TABLE 10 | Comparison of comparison of FRP and VPM with SPECT images.

	Sensitivity rate with false positive rates on conventional models					
Models	**1/8 (%)**	**1/4 (%)**	**1/2 (%)**	**2 (%)**	**4 (%)**	**VPM (%)**
KNN	94.77	94.03	93.05	94.29	93.63	67.70
CNN	95.06	95.36	95.79	95.92	95.56	77.77
ANN	97.78	96.47	96.98	96.03	96.27	89.03
Proposed model	98.00	98.11	98.47	96.12	96.18	94.08
	Sensitivity rates with false positive rates on existing works					
Methods	**1/8 (%)**	**1/4 (%)**	**1/2 (%)**	**2 (%)**	**4 (%)**	**VPM (%)**
[46]	94.17	95.23	95.46	95.75	95.07	80.60
[47]	95.37	96.59	96.80	96.17	96.00	90.80
[48]	93.42	94.89	97.39	96.89	97.45	94.00
Proposed model	96.07	96.19	98.36	97.09	98.07	99.93

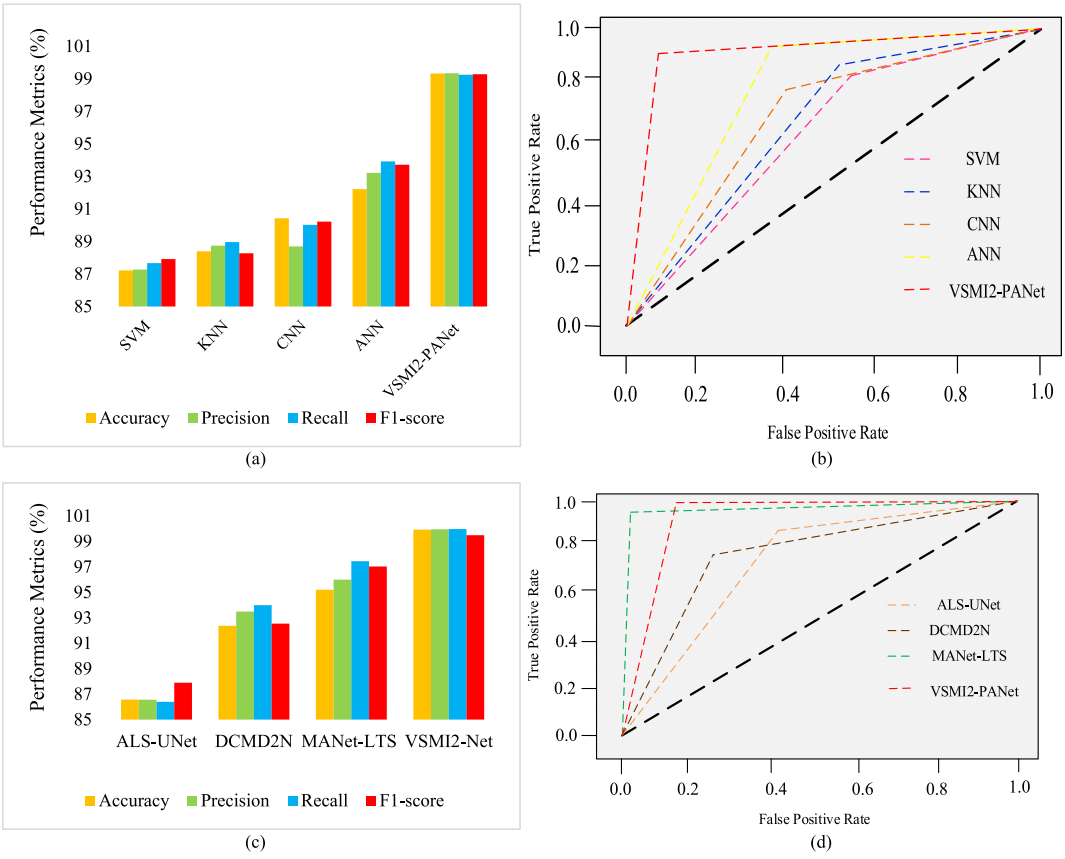


FIGURE 5 | (a) Comparison of proposed versus conventional models using SPECT images. (b) Comparison of proposed versus existing works using SPECT images. (c) AUC comparison on proposed versus existing models using SPECT images. (d) AUC comparison of proposed versus existing models using SEPCT images.

images are provided as input at the LWT for the comparison purpose, the performance degrades due to its increased illumination nature and higher cost. Overall, the adoption of LWT as a backbone model with multi modal imaging techniques enhances the detection capability of the lung tumour.

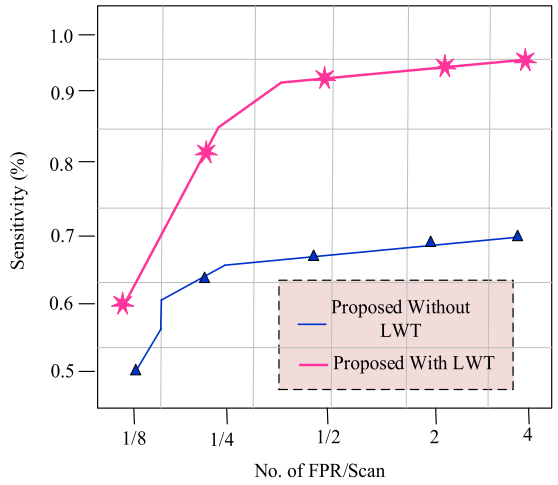


FIGURE 6 | No of FPR/scan versus sensitivity (%).

7 | Discussion

In the real-world situations depicted in Figures 7–9, the proposed model illustrates its ability to handle complicated and difficult conditions typically seen in medical imaging, notably in the detection of lung cancers. Tumour morphologies differ significantly in form, size and texture, making precise segmentation challenging. Tumours can sometimes look concealed due to noise in the imaging data or a lack of contrast between the tumour and the surrounding tissue. Traditional machine learning models and simpler deep learning architectures struggle with these complexities, frequently failing to precisely define tumour borders. However, our model's innovative architecture efficiently overcomes these issues. The versatile scale-malleable image integration ($VSMI^2$) module is critical in this reard because it is designed to process and integrate multimodal images from CT, MRI and SPECT, all of which provide complementary spatial and spectral information. The ability to use

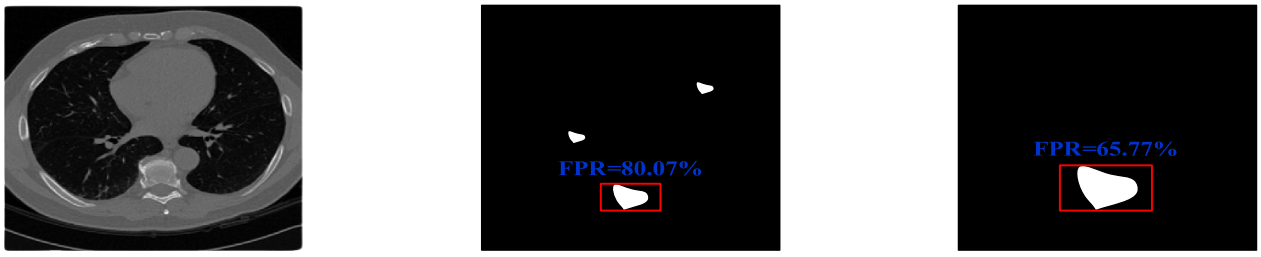


FIGURE 7 | (a) The illustration of original, (b) without LWT and (c) with LWT from CT images.

FIGURE 8 | (a) The illustration of original, (b) without LWT and (c) with LWT from MRI images.

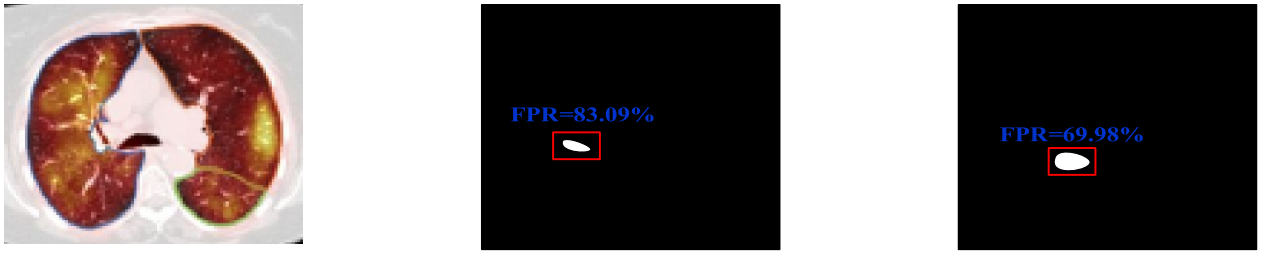


FIGURE 9 | (a) The illustration of original, (b) without LWT and (c) with LWT from SPECT images.

multi-modal data allows the model to catch detailed characteristics that may not be completely evident in a single modality, resulting in more precise and accurate tumour segmentation, especially in the face of noise or unclear borders.

The patch-wise attention network (PANet), another important component of the model, improves segmentation by concentrating on specific regions of interest in the picture. This patch-wise attention technique enables the model to direct greater computing resources and attention to places where tumours are most likely to be found while still evaluating the image's overall context. This approach is very useful when dealing with complicated instances in which tumours have irregular forms or are located in difficult areas of the lung. By focusing on these crucial locations, PANet guarantees that even the tiniest or most irregular tumour patches are correctly divided. Furthermore, the light wave transformer (LWT) built into PANet provides an improved method for dealing with noise and improving the sensitivity of the segmentation process, making it ideal for pictures that are less clear owing to low resolution or poor quality.

Compared to other methodologies, the suggested model offers a substantial benefit. Traditional machine learning algorithms, including basic convolutional neural networks (CNNs), are unable to capture the fine-grained information that multi-modal imagery and attention-based models such as PANet provide. Our model's emphasis on retaining spatial context with VSMI^2 and improving region-specific attention with PANet leads to better performance in difficult clinical circumstances. The therapeutic implications are significant, as this degree of accuracy in tumour segmentation allows for earlier and more precise therapies, potentially improving patient outcomes and making the model extremely useful in medical practice.

8 | Conclusion

Our study addressed the critical issue of lung cancer detection and prediction which is a global health concern with a high mortality rate. We proposed a novel approach named as versatile scale malleable image integration and patch-wise attention network (VSMI^2-PANet), which leverages the power of machine learning and deep learning techniques. By integrating information from multiple imaging modalities including computed tomography (CT), magnetic resonance imaging (MRI) and single photon emission computed tomography (SPECT), our model offered a comprehensive solution for early lung tumour detection. VSMI^2-PANet employed a multi-step process, started with image cropping and scale malleable convolution layer (SMCL) modules to extract meaningful patches and preserve spatial information. The patch-wise attention network (PANet) further refined the information by analysing image patch characteristics, generating high-resolution spatial attention maps pinpointing suspicious tumour locations. These attention maps were then fed into a backbone module that utilised the light wave transformer for the classification of lung tumours into three classes: normal, benign and malignant. Additionally, the model incorporated SPECT images to capture variations accurately for the segmentation task.

Our work was implemented and validated using MATLAB simulation with meticulous system and simulation configurations. The performance evaluations were based on metrics such as accuracy, precision, recall, *F*1-score and AUC curve. The final results clearly demonstrated the superiority of our proposed model over existing methods. This innovative approach worked effectively for early lung cancer detection and contributed to the ongoing efforts to reduce the global burden of this deadly disease.

In the future, we plan to investigate the potential of transfer learning by pre-training on a large dataset, possibly from a related field like general medical imaging, and fine-tuning the model on lung-specific data. This can be help in cases where limited lung cancer data is available. Furthermore, we aim to explore advanced cross-modality feature learning in term of segmentation, reliable mutual distillation techniques and capsule networks to further improve the performance of the model while analysing complicated patterns [49–53].

Acknowledgements

The work of Muhammad Faheem is supported by the VTT Technical Research Centre of Finland and the work of Nayef Alqahtani is supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia (Grant KFU251882).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The dataset is openly available and explained in the manuscript.

References

1. S. Alagarsamy, R. Raja Subramanian, T. Shree, S. Kannan, M. Balasubramanian, and V. Govindaraj, "Prediction of Lung Cancer Using Meta-Heuristic Based Optimization Technique: Crow Search Technique," in *2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)* (2021), 186–191.

2. M. Mamun, A. Farjana, M. A. Mamun, and M. S. Ahammed, "Lung Cancer Prediction Model Using Ensemble Learning Techniques and a Systematic Review Analysis," in *2022 IEEE World AI IoT Congress (AIIoT)* (2022), 187–193.

3. C. Venkatesan, D. Balamurugan, T. Thamaraimanalan, and M. Ramkumar, "Efficient Machine Learning Technique for Tumor Classification Based on Gene Expression Data," in *2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Vol. 1 (2022), 1982–1986, https://doi.org/10.1109/icaccs54159.2022.9785294.

4. L. Chang, J. Wu, N. Moustafa, A. K. Bashir, and K. Yu, "AI-Driven Synthetic Biology for Non-Small Cell Lung Cancer Drug Effectiveness-Cost Analysis in Intelligent Assisted Medical Systems," *IEEE Journal of Biomedical and Health Informatics* 26, no. 10 (2021): 5055–5066, https://doi.org/10.1109/jbhi.2021.3133455.

5. J. Li, Y. Tao, and T. Cai, "Predicting Lung Cancers Using Epidemiological Data: A Generative-Discriminative Framework," *IEEE/CAA Journal of Automatica Sinica* 8, no. 5 (2021): 1067–1078, https://doi.org/10.1109/jas.2021.1003910.

6. U. N. Wisesty, A. Purwarianti, A. Pancoro, et al., "Join Classifier of Type and Index Mutation on Lung Cancer DNA Using Sequential

Labeling Model," *IEEE Access* 10 (2022): 9004–9021, https://doi.org/10.1109/access.2022.3142925.

7. L. Zhao, J. Qian, F. Tian, et al., "A Weighted Discriminative Extreme Learning Machine Design for Lung Cancer Detection by an Electronic Nose System," *IEEE Transactions on Instrumentation and Measurement* 70 (2021): 1–9, https://doi.org/10.1109/tim.2021.3084312.

8. M. Khushi, K. Shaukat, T. M. Alam, et al., "A Comparative Performance Analysis of Data Resampling Methods on Imbalance Medical Data," *IEEE Access* 9 (2021): 109960–109975, https://doi.org/10.1109/access.2021.3102399.

9. R. Baghbani, M. B. Shadmehr, M. Ashoorirad, S. F. Molaeezadeh, and M. H. Moradi, "Bioimpedance Spectroscopy Measurement and Classification of Lung Tissue to Identify Pulmonary Nodules," *IEEE Transactions on Instrumentation and Measurement* 70 (2021): 1–7, https://doi.org/10.1109/tim.2021.3105241.

10. R. Kumar Sachdeva, T. Garg, G. S. Khaira, D. Mitrav, and R. Ahuja, "A Systematic Method for Lung Cancer Classification," in *2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)* (2022), 1–5.

11. V. G. Biradar, P. K. Pareek, V. S, and N. P, "Lung Cancer Detection and Classification Using 2D Convolutional Neural Network," in *2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon)* (2022), 1–5.

12. Z. Zhou, F. Gou, Y. Tan, and J. Wu, "A Cascaded Multi-Stage Framework for Automatic Detection and Segmentation of Pulmonary Nodules in Developing Countries," *IEEE Journal of Biomedical and Health Informatics* 26, no. 11 (2022): 5619–5630, https://doi.org/10.1109/jbhi.2022.3198509.

13. A. Rehman, M. Kashif, I. Abunadi, and N. Ayesha, "Lung Cancer Detection and Classification From Chest CT Scans Using Machine Learning Techniques," in *2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA)* (2021), 101–104.

14. N. Cherukuri, N. R. Bethapudi, V. S. Thotakura, P. Chitturi, C. Z. Basha, and R. M. Mummidi, "Deep Learning for Lung Cancer Prediction Using NSCLS Patients CT Information," in *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)* (2021), 325–330.

15. M. Praveena, A. Ravi, T. Srikanth, B. Praveen, B. S. Krishna, and A. K. Mallik, "Lung Cancer Detection Using Deep Learning Approach CNN," in *2022 7th International Conference on Communication and Electronics Systems (ICCES)* (2022), 1418–1423.

16. D. Florensa, P. Godoy, J. Mateo, et al., "The Use of Multiple Correspondence Analysis to Explore Associations Between Categories of Qualitative Variables and Cancer Incidence," *IEEE Journal of Biomedical and Health Informatics* 25, no. 9 (2021): 3659–3667, https://doi.org/10.1109/jbhi.2021.3073605.

17. Y. Chen, Y. Wang, F. Hu, L. Feng, T. Zhou, and C. Zheng, "LDNNET: Towards Robust Classification of Lung Nodule and Cancer Using Lung Dense Neural Network," *IEEE Access* 9 (2021): 50301–50320, https://doi.org/10.1109/access.2021.3068896.

18. F. Silva, T. Pereira, J. Morgado, et al., "EGFR Assessment in Lung Cancer CT Images: Analysis of Local and Holistic Regions of Interest Using Deep Unsupervised Transfer Learning," *IEEE Access* 9 (2021): 58667–58676, https://doi.org/10.1109/access.2021.3070701.

19. Y. Wu, J. Ma, X. Huang, S. Ling, and S. W. Su, "DeepMMSA: A Novel Multimodal Deep Learning Method for Non-Small Cell Lung Cancer Survival Analysis," in *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (2021), 1468–1472.

20. D. Kadia, M. Z. Alom, R. Burada, T. V. Nguyen, and V. K. Asari, "R2U3D: Recurrent Residual 3D U-Net for Lung Segmentation," *IEEE Access* 9 (2021): 88835–88843, https://doi.org/10.1109/access.2021.3089704.

21. L. Liu, K. Fan, and M. Yang, "Federated Learning: A Deep Learning Model Based on ResNet18 Dual Path for Lung Nodule Detection," *Multimedia Tools and Applications* 82, no. 11 (2023): 17437–17450, https://doi.org/10.1007/s11042-022-14107-0.

22. L. Chen, K. Liu, H. Shen, et al., "Multimodality Attention-Guided 3-D Detection of Nonsmall Cell Lung Cancer in 18F-FDG PET/CT Images," *IEEE Transactions on Radiation and Plasma Medical Sciences* 6, no. 4 (2021): 421–432, https://doi.org/10.1109/trpms.2021.3072064.

23. M. Tortora, E. Cordelli, R. Sicilia, et al., "RadioPathomics: Multimodal Learning in Non-Small Cell Lung Cancer for Adaptive Radiotherapy," *IEEE Access* 11 (2022): 47563–47578, https://doi.org/10.1109/access.2023.3275126.

24. C. Bian, C. Yuan, K. Ma, S. Yu, D. Wei, and Y. Zheng, "Domain Adaptation Meets Zero-Shot Learning: An Annotation-Efficient Approach to Multi-Modality Medical Image Segmentation," *IEEE Transactions on Medical Imaging* 41, no. 5 (2021): 1043–1056, https://doi.org/10.1109/tmi.2021.3131245.

25. J. Li, H. Chen, Y. Li, Y. Peng, J. Sun, and P. Pan, "Cross-Modality Synthesis Aiding Lung Tumor Segmentation on Multi-Modal MRI Images," *Biomedical Signal Processing and Control* 76 (2022): 103655, https://doi.org/10.1016/j.bspc.2022.103655.

26. J. R. Barrett and T. Viana, "EMM-LC Fusion: Enhanced Multimodal Fusion for Lung Cancer Classification," *AI* 3, no. 3 (2022): 659–682, https://doi.org/10.3390/ai3030038.

27. C. M. Caruso, V. Guarrasi, E. Cordelli, et al., "A Multimodal Ensemble Driven by Multiobjective Optimisation to Predict Overall Survival in Non-Small-Cell Lung Cancer," *Journal of Imaging* 8, no. 11 (2022): 298, https://doi.org/10.3390/jimaging8110298.

28. Z. Li, S. Wang, H. Yu, et al., "A Novel Deep Learning Framework Based Mask-Guided Attention Mechanism for Distant Metastasis Prediction of Lung Cancer," *IEEE Transactions on Emerging Topics in Computational Intelligence* 7, no. 2 (2023): 330–341, https://doi.org/10.1109/tetci.2022.3171311.

29. A. Syed Musthafa, K. Sankar, T. Benil, and Y. N. Rao, "A Hybrid Machine Learning Technique for Early Prediction of Lung Nodules From Medical Images Using a Learning-Based Neural Network Classifier," *Concurrency and Computation: Practice and Experience* 35, no. 3 (2023): e7488, https://doi.org/10.1002/cpe.7488.

30. M. A. Balcı, L. M. Batrancea, Ö. Akgüller, and A. Nichita, "A Series-Based Deep Learning Approach to Lung Nodule Image Classification," *Cancers* 15, no. 3 (2023): 843, https://doi.org/10.3390/cancers15030843.

31. K. Suzuki, Y. Otsuka, Y. Nomura, K. K. Kumamaru, R. Kuwatsuru, and S. Aoki, "Development and Validation of a Modified Three-Dimensional U-Net Deep-Learning Model for Automated Detection of Lung Nodules on Chest CT Images From the Lung Image Database Consortium and Japanese Datasets," *Academic Radiology* 29 (2022): S11–S17, https://doi.org/10.1016/j.acra.2020.07.030.

32. D. Bhattacharyya, N. Thirupathi Rao, E. S. N. Joshua, and Y. C. Hu, "A Bi-Directional Deep Learning Architecture for Lung Nodule Semantic Segmentation," *Visual Computer* 39, no. 11 (2023): 5245–5261, https://doi.org/10.1007/s00371-022-02657-1.

33. V. K. Gugulothu and S. Balaji, "An Early Prediction and Classification of Lung Nodule Diagnosis on CT Images Based on Hybrid Deep Learning Techniques," *Multimedia Tools and Applications* 83, no. 1 (2024): 1041–1061.

34. J. Ruan, Y. Meng, F. Zhao, H. Gu, L. He, and X. Gong, "Development of Deep Learning-Based Automatic Scan Range Setting Model for Lung Cancer Screening Low-Dose CT Imaging," *Academic Radiology* 29, no. 10 (2022): 1541–1551, https://doi.org/10.1016/j.acra.2021.12.001.

35. I. Ahmed, A. Chehri, G. Jeon, and F. Piccialli, "Automated Pulmonary Nodule Classification and Detection Using Deep Learning Architectures," *IEEE/ACM Transactions on Computational Biology and*

Bioinformatics 20, no. 4 (2022): 2445–2456, https://doi.org/10.1109/tcbb.2022.3192139.

36. X. Cui, S. Zheng, M. A. Heuvelmans, et al., "Performance of a Deep Learning-Based Lung Nodule Detection System as an Alternative Reader in a Chinese Lung Cancer Screening Program," *European Journal of Radiology* 146 (2022): 110068, https://doi.org/10.1016/j.ejrad.2021.110068.

37. R. Manickavasagam, S. Selvan, and M. Selvan, "CAD System for Lung Nodule Detection Using Deep Learning With CNN," *Medical & Biological Engineering & Computing* 60, no. 1 (2022): 221–228, https://doi.org/10.1007/s11517-021-02462-3.

38. H. Huang, R. Wu, Y. Li, and C. Peng, "Self-Supervised Transfer Learning Based on Domain Adaptation for Benign-Malignant Lung Nodule Classification on Thoracic CT," *IEEE Journal of Biomedical and Health Informatics* 26, no. 8 (2022): 3860–3871, https://doi.org/10.1109/jbhi.2022.3171851.

39. G. Kasinathan and S. Jayakumar, "Cloud-Based Lung Tumor Detection and Stage Classification Using Deep Learning Techniques," *BioMed Research International* 2022, no. 1 (2022), https://doi.org/10.1155/2022/4185835.

40. I. Ali, M. Muzammil, I. U. Haq, A. A. Khaliq, and S. Abdullah, "Deep Feature Selection and Decision Level Fusion for Lungs Nodule Classification," *IEEE Access* 9 (2021): 18962–18973, https://doi.org/10.1109/access.2021.3054735.

41. P. J. Muhammad Ali, "Investigating the Impact of Min-Max Data Normalization on the Regression Performance of K-Nearest Neighbor With Different Similarity Measurements," *ARO-The Scientific Journal of Koya University* 10, no. 1 (2022): 85–91, https://doi.org/10.14500/aro.10955.

42. Y. Jiang, B. Wronski, B. Mildenhall, J. T. Barron, Z. Wang, and T. Xue, "Fast and High-Quality Image Denoising Via Malleable Convolutions," in *European Conference on Computer Vision* (2022), 1–5.

43. S. Lee, J. Lee, B. Kim, E. Yi, and J. Kim, "Patch-Wise Attention Network for Monocular Depth Estimation," in *AAAI Conference on Artificial Intelligence* (2021), 23–32.

44. M. Gwak, J. Cha, H. Yoon, D. Kang, and D. An, "Lightweight Transformer Model for Mobile Application Classification," *Sensors (Basel, Switzerland)* 24, no. 2 (2024): 564, https://doi.org/10.3390/s24020564.

45. X. Li, L. Song, S. Liu, and C. Lv, "A Lung Nodule Detection Method Based on Faster R-CNN," *Computer Methods and Programs in Biomedicine* 177 (2019): 237–243.

46. R. Gao, Q. Lin, Z. Man, and Y. Cao, "Automatic Lesion Segmentation of Metastases in SPECT Images Using U-Net-Based Model," in *Other Conferences* (2022), 44–49.

47. H. Xie, Z. Liu, L. Shi, et al., "Segmentation-Free PVC for Cardiac SPECT Using a Densely-Connected Multi-Dimensional Dynamic Network," *IEEE Transactions on Medical Imaging* 42, no. 5 (2022): 1325–1336, https://doi.org/10.1109/tmi.2022.3226604.

48. S. Saeku, N. Noipinit, K. Khamwan, and P. Siricharoen, "Liver and Tumor Segmentation in Selective Internal Radiation Therapy 99mTc-MAA SPECT/CT Images Using MANet and Histogram Adjustment," in *2022 3rd Asia Symposium on Signal Processing (ASSP)* (2022), 62–66.

49. D. Zhang, G. Huang, Q. Zhang, J. Han, J. Han, and Y. Yu, "Cross-Modality Deep Feature Learning for Brain Tumor Segmentation," *Pattern Recognition* 110 (2021): 107562, https://doi.org/10.1016/j.patcog.2020.107562.

50. C. Fang, Q. Wang, L. Cheng, et al., "Reliable Mutual Distillation for Medical Image Segmentation Under Imperfect Annotations," *IEEE Transactions on Medical Imaging* 42, no. 6 (2023): 1720–1734, https://doi.org/10.1109/tmi.2023.3237183.

51. Y. Liu, D. Zhang, Q. Zhang, and J. Han, "Part-Object Relational Visual Saliency," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44 (2021): 3688–3704.

52. Y. Liu, L. Zhou, G. Wu, S. Xu, and J. Han, "TCGNet: Type-Correlation Guidance for Salient Object Detection," *IEEE Transactions on Intelligent Transportation Systems* 25, no. 7 (July 2024): 6633–6644, https://doi.org/10.1109/TITS.2023.3342811.

53. Y. Liu, D. Cheng, D. Zhang, S. Xu, and J. Han, "Capsule Networks With Residual Pose Routing," *IEEE Transactions on Neural Networks and Learning Systems* 36, no. 2 (2024): 2648–2661, https://doi.org/10.1109/tnnls.2023.3347722.