Research Article

Self-Supervised Graph Vision Transformer Network For Early Pulmonary Nodule Detection And Malignancy Classification From Ct Imaging

Koventhan G M[1], Dr. R. Ramkumar[2]

[1]Research Scholar, Department of Computer Science, Sasurie College of Arts & Science, Naduppatti, Erode, Tamil Nadu, Erode.
[2]Principal, Shree Venkateshwara Arts and Science (Co-education) College, Othakuthirai, Gobichettipalayam, Erode District.

Abstract

Early detection of pulmonary nodules is very important in enhancing survival rates among lung cancer patients. Nevertheless, the small nodules are difficult to identify and classify in computed tomography (CT) images because of their insignificant appearance, change in size and shape, and the limited amount of large annotated medical data. To mitigate these shortcomings, this paper suggests a Self-Supervised Graph Vision Transformer Network (SS-GVTNet) in detecting and classifying pulmonary nodules and malignancies through CT scanning. The suggested framework combines self-supervised representation learning with the graph-based spatial modeling and transformer-based global feature extraction. First, self-supervised pretraining module trains powerful feature representations using massive amounts of unlabeled CT scans, which removes reliance on manual labels. A Self-Supervised Graph Vision Transformer (SGVT) backbone is then used to capture cross-slice contextual dependencies of CTs, and successful modeling of complex anatomy patterns is possible. To further improve the spatial cognition, a Enhanced Granular Neural Network (EGNN) module is built to form connectivity relations among possible nodule areas to enable the neural net to learn structural associations and contextual interactions in lung tissues. The hybrid architecture enhances the local feature removal and the global contextual reasoning. Empirical testing of publicly available lung CT data proves that the suggested model has higher detection accuracy, sensitivity, and specificity than the existing convolutional neural networks and mixed deep learning methods. Moreover, the framework offers credible classification of malignancy, which help clinicians in early diagnosis and treatment planning. The proposed solution is a scalable and efficient computer-aided lung cancer screening system that involves the integration of self-supervised learning, the graph-based modeling of features, and transformer architecture.

Keywords: Pulmonary nodule detection, self-supervised learning, granular neural networks, vision transformer, lung cancer diagnosis, CT image analysis, deep learning.

1. Introduction

Early detection of pulmonary nodules plays a crucial role in reducing mortality associated with Lung Cancer, which remains one of the leading causes of cancer-related deaths worldwide [1-3]. Pulmonary nodules are small abnormal growths in lung tissue that may represent benign lesions or early-stage malignant tumors. Detecting these nodules at an early stage significantly improves treatment outcomes and survival rates. Medical imaging techniques such as Computed Tomography have become essential diagnostic tools because they provide high-resolution images of lung structures, enabling clinicians to identify subtle abnormalities that may indicate the presence of pulmonary nodules [4-5]. Despite the effectiveness of CT imaging, manual interpretation of CT scans is time-consuming and highly dependent on radiologists' expertise. Small nodules often exhibit subtle variations in size, shape, and texture, which can make accurate detection and classification challenging [6]. Furthermore, the increasing number of CT examinations in clinical practice has created a need for automated and intelligent diagnostic systems that can assist clinicians in identifying suspicious nodules quickly and accurately [7]. Consequently, artificial intelligence and deep learning approaches have gained significant attention for developing computer-aided diagnosis systems that enhance diagnostic accuracy and efficiency.

Recent advances in Deep Learning have enabled the development of powerful models for medical image analysis [8-10]. In particular, architectures based on Computer Vision, such as convolutional neural networks and transformer-based models, have demonstrated remarkable performance in detecting complex patterns from medical imaging data. Among these approaches, the Vision Transformer has shown strong capability in capturing long-range dependencies within images by utilizing self-attention mechanisms [11]. However, conventional transformer models often require large annotated datasets for effective training, which can be difficult to obtain in medical domains due to limited labeled data and the need for expert annotations. To address this limitation, Self-Supervised Learning has emerged as an effective paradigm that allows models to learn meaningful feature representations from unlabeled data. By generating supervisory signals directly from the data itself, self-supervised approaches reduce the dependence on large labeled datasets while improving model generalization. Additionally, representing medical images using granular structures can capture spatial relationships between different regions of the lung. In this context, Granular Neural Networks provide an effective mechanism for modeling complex relationships between nodules and surrounding tissues.

Motivated by these advancements, this research proposes a Self-Supervised Graph Vision Transformer Network for early pulmonary nodule detection and malignancy classification from CT imaging. The proposed framework integrates self-supervised learning with graph-based representations and transformer architectures to capture both global contextual information and local structural relationships in lung CT images. By combining graph modeling and transformer attention mechanisms, the system aims to enhance feature extraction, improve detection accuracy, and enable reliable classification of benign and malignant nodules. The proposed method is designed to support clinicians by providing an automated, accurate, and efficient diagnostic tool for early pulmonary nodule detection. Ultimately, such intelligent systems can contribute to earlier diagnosis of lung cancer, improved clinical decision-making, and better patient outcomes.

The rest of the research work is structured as follows as, section 2 reviews the some of the recent techniques for the detection of pulmonary nodule. section 3 presents the process of the proposed methodology. section 4 provides the results and discussion. section 5 deals with the conclusion and future work.

2. Literature Review

This section reviews the some of the recent techniques for the detection of pulmonary nodule using advanced machine learning and deep learning techniques.

Wang et al [12] introduced a preoperative and intraoperative methods for localizing lung nodules before minimally invasive resection. Methods for preoperative localization include CT-guided hook-wire positioning, coil positioning, or dye injection and radionuclide location Methods for intraoperative localization include intraoperative ultrasound localization and tactile pressure-sensing localization. After the localization of pulmonary nodules under the guidance of CT patients need to restrict their activities; otherwise, it is easy for the nodules to move, causing the operation to fail, and may also cause complications such as pneumothorax, puncture site pain, and pulmonary parenchymal bleeding. In the past, we injected melamine dye under the guidance of electromagnetic navigation bronchoscope to locate lung nodules. The purpose of this case is introducing a new method for accurately localizing and resecting pulmonary nodules by injecting indocyanine green (ICG) under the guidance of electromagnetic navigation bronchoscope and the resection of small pulmonary nodules under the fluoroscope.

Nasrullah et al [13] a novel deep learning-based model with multiple strategies is proposed for the precise diagnosis of the malignant nodules. Due to the recent achievements of deep convolutional neural networks (CNN) in image analysis, we have used two deep three-dimensional (3D) customized mixed link network (CMixNet) architectures for lung nodule detection and classification, respectively. Nodule detections were performed through faster R-CNN on efficiently-learned features from CMixNet and U-Net like encoder–decoder architecture. Classification of the nodules was performed through a gradient boosting machine (GBM) on the learned features from the designed 3D CMixNet structure. To reduce false positives and misdiagnosis results due to different types of errors, the final decision was performed in connection with physiological symptoms and clinical biomarkers. With the advent of the internet of things (IoT) and electro-medical technology, wireless body area networks (WBANs) provide continuous monitoring of patients, which helps in diagnosis of chronic diseases—especially metastatic cancers. The deep learning model for nodules' detection and classification, combined with clinical factors, helps in the reduction of misdiagnosis and false positive (FP) results in early-stage lung cancer diagnosis. The proposed system was evaluated on LIDC-IDRI datasets in the form of sensitivity (94%) and specificity (91%), and better results were obatined compared to the existing methods.

Shariaty et al [14] developed a Computer-Aided Detection systems for the diagnosis of lung cancer has been studied - including preprocessing and segmentation methods, and data analysis techniques. The main goal was to investigate the latest technology for the development of computational diagnostic tools, so as to assist with the acquisition, processing, and analysis of the medical imagery data. However, there are aspects that still require further attention, such as to increase the sensitivity, reduce false positives, and optimize the detection of each type of nodule, even for differing size and shape.

Sori et al [15] proposed a deep CNN architecture which differs from those traditionally used in computer vision to solve this problem. First, the suspicious nodules are generated with the modified version of U-Net and then the generated nodules become an input data for our model. The proposed model is a multi-path CNN which exploits both local features as well as more global contextual features simultaneously to automatically detect lung cancer. To this end, the model used three paths, each path employed different receptive field size which helps to model distant dependencies (short and long-range dependencies of the neighboring pixels). Then, to further upgrade our model performance, we concatenate features from the three paths. This balance the receptive field size effect and makes our model more adaptable to the variability of shape, size, and contextual information among nodules. Finally, we also introduce a retraining phase system that permits us to tackle difficulties related to the imbalance of image labels. Experimental results on Kaggle Data Science Bowl 2017 challenge shows that our model is better adaptable to the described inconsistency among nodules size and shape, and also obtained better detection results compared to the recently published state of the art methods.

Sang et al [16] presented a deep learning-based framework for automated lung cancer detection. The proposed framework works in multiple stages on 3D lung CT scan images to detect and determine the malignancy of the nodules. Considering 3D nature of lung CT data and the compactness of mixed link network (MixNet), two deep 3D faster R-CNN and U-Net encoder-decoder with MixNet were designed to detect and learn the features of the lung nodule, respectively. For the classification of the nodules, the gradient boosting machine (GBM) with 3D MixNet was proposed. The proposed system was tested with manually draw radiologist contours on 1200 images obtained from LIDC-IDRI including 3250 nodules by using statistical measures. LIDC-IDRI comprises of equal number of benign and malignant lung nodules. The proposed system was evaluated on this data set in the form of sensitivity (94%), specificity (90%), area under the receiver operating curve (0.99) and obtained better results compared to the existing methods.

Zhang et al [17] suggested an efficient identification system for lung nodules based on Multi-Scene Deep Learning Framework (MSDLF) by the vesselness filter. A four-channel CNN model is designed to enhance the radiologist's knowledge in the detection of four-stage nodules by integrating two image Scenes. This model can be applied in two different classes. The results show that the Multi-Scene Deep Learning Framework (MSDLF) is efficient for increasing the accuracy and significantly reducing false positives in an enormous amount of image data in the detection of lung nodules.

UrRehman et al [18] introduced a bespoke convolutional neural network (CNN) with a dual attention mechanism was created, which was especially crafted to concentrate on the most important elements in images of lung nodules. The CNN model extracts informative features from the images, while the attention module incorporates both channel attention and spatial attention mechanisms to selectively highlight significant features. After the attention module, global average pooling is applied to summarize the spatial information. To evaluate the performance of the proposed model, extensive experiments were conducted using benchmark dataset of lung nodules. The results of these experiments demonstrated that our model surpasses recent models and achieves state-of-the-art accuracy in lung nodule detection and classification tasks.

Srivastava et al [19] developed a Hybridized Faster R-CNN (HFRCNN) to identify lung cancer at an early stage. Among the numerous uses for which faster R-CNN has been put to good use is identifying critical entities in medical imagery, such as MRIs and CT scans. Many research investigations in recent years have examined the use of various techniques to detect lung nodules (possible indicators of lung cancer) in scanned images, which may help in the early identification of lung cancer. One such model is HFRCNN, a two-stage, region-based entity detector. It begins by generating a collection of proposed regions, which are subsequently classified and refined with the aid of a convolutional neural network (CNN). A distinct dataset is used in the model’s training process, producing valuable outcomes. More than a 97% detection accuracy was achieved with the suggested model, making it far more accurate than several previously announced methods.

Mothkur et al [20] developed a lightweight deep neural net (DNN) designs for image processing. The proposed lightweight deep neural model outperforms with an accuracy of 85.21% with reasonable specificity and sensitivity trade-off. The proposed work is based on binary classification networks such as vanilla 2D CNN, 2D SqueezeNet, and 2D MobileNet to recognize lung cancer in patient CT images with and without early-stage lung cancer.

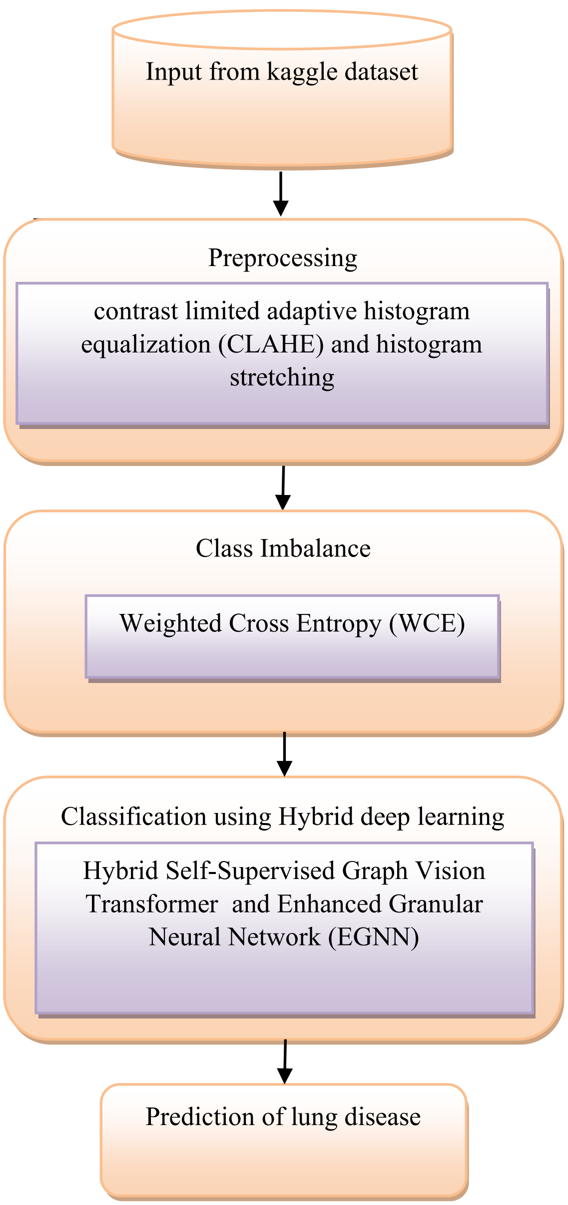
Cao et al [21] developed a robust nodule detection method. In this study, we propose a two-stage convolutional neural networks (TSCNN) for lung nodule detection. The first stage based on the improved U-Net segmentation network is to establish an initial detection of lung nodules. During this stage, in order to obtain a high recall rate without introducing excessive false positive nodules, we propose a new sampling strategy for training. Simultaneously, a two-phase prediction method is also proposed in this stage. The second stage in the TSCNN architecture based on the proposed dual pooling structure is built into three 3D-CNN classification networks for false positive reduction. Since the network training requires a significant amount of training data, we designed a random mask as the data augmentation method in this study. Furthermore, we have improved the generalization ability of the false positive reduction model by means of ensemble learning. We verified the proposed architecture on the LUNA dataset in our experiments, which showed that the proposed TSCNN architecture did obtain competitive detection performance.

Oumlaz et al [22] presented a hybrid model proficiently handles the complexities of 3D CT images, ensuring precise and reliable lung nodule detection. The algorithm processes CT scans using short slice grouping (SSG) and long slice grouping (LSG) techniques to extract critical features from each slice, culminating in the generation of nodule probabilities and the identification of potential nodular regions. Incorporating shapley additive explanations (SHAP) analysis further enhances model interpretability by highlighting the contributory features. Our extensive experimentation demonstrated a significant improvement in diagnostic accuracy, with training accuracy increasing from 0.9126 to 0.9817. This advancement not only reflects the model’s efficient learning curve but also its high proficiency in accurately classifying a majority of training samples. Given its high accuracy, interpretability, and consistent reduction in training loss, ARSGNet holds substantial potential as a groundbreaking tool for early lung cancer detection and diagnosis.

3. Proposed Methodology

This paper suggests a Self-Supervised Graph Vision Transformer Network (SS-GVTNet) in detecting and classifying pulmonary nodules and malignancies through CT scanning. The suggested framework combines self-supervised representation learning with the graph-based spatial modeling and transformer-based global feature extraction. First, self-supervised pretraining module trains powerful feature representations using massive amounts of unlabeled CT scans, which removes reliance on manual labels. A Vision Transformer (ViT) backbone is then used to capture cross-slice contextual dependencies of CTs, and successful modeling of complex anatomy patterns is possible. To further improve the spatial cognition, a Granular Neural Network (GNN) module is built to form connectivity relations among possible nodule areas to enable the neural net to learn structural associations and contextual interactions in lung tissues. The hybrid architecture enhances the local feature removal and the global contextual reasoning. Empirical testing of publicly available lung CT data proves that the suggested model has higher detection accuracy, sensitivity, and specificity than the existing convolutional neural networks and mixed deep learning methods. Moreover, the framework offers credible classification of malignancy, which help clinicians in early diagnosis and treatment planning. The figure 1. shows the process of the proposed methodology.

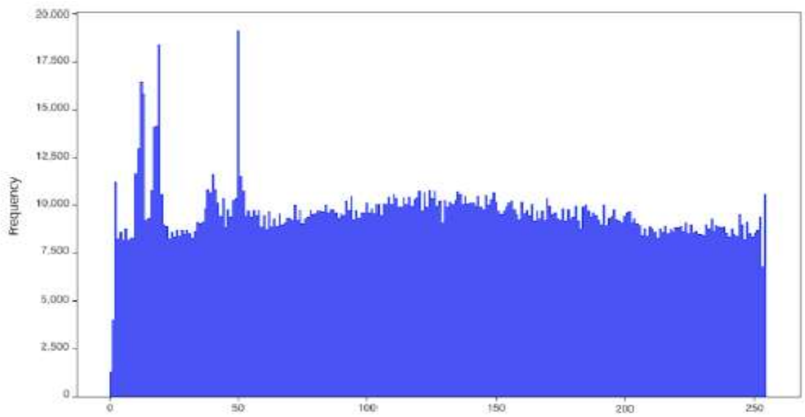
Figure 1. the process of the proposed methodology



3.1. Pre-processing

In this research work introduced a contrast limited adaptive histogram equalization (CLAHE) and histogram stretching. CLAHE enhances local contrast without over-amplifying noise, while histogram stretching evenly spreads out the pixel values across both datasets. Figures 2 highlight the uniformity of the pixel intensities for both the training and testing datasets after preprocessing. The training dataset shows a more balanced distribution with minor peaks, especially around intensity values below 50 and near 200. The testing dataset follows a similar pattern, with a more even spread of pixel intensities. This alignment minimizes domain shift, improving the models ability to generalize to unseen data.

Figure 2. Pixel intensity distribution of the training dataset after preprocessing.



This preprocessing effectively reduced intensity variations and standardized the images, which is crucial for stabilizing the CNN training process. The improved uniformity across datasets supports consistent feature extraction, which is expected to enhance the final model's performance.

3.2. Class Imbalance in Medical Diagnostic Datasets

In the domain of medical imaging analysis, one encounters the significant hurdle of class imbalance within diagnostic datasets. The presence of such an imbalance can skew a model's learning, leading it to favor the more prevalent class and, thus, compromise its diagnostic performance. Consider the standard loss function used in classification, cross-entropy loss, for which a single prediction is given by

$$\mathcal{L}_{\mathrm{CE}}(\mathrm{x_i}) = -\mathrm{y_i}\log\mathrm{f}(\mathrm{x_i}) - (1-\mathrm{y_i})\log(1-\mathrm{f}(\mathrm{x_i})) \qquad (1)$$

where CE is cross-entropy, yi is the true label, $\mathrm{f}(\mathrm{x_i})$ is the predicted probability of the positive class, and $(1-\mathrm{y_i})$ is the predicted probability of the negative class. For the entire training dataset D with N total examples, the average cross-entropy loss can be expressed as

$$\mathcal{L}_{\mathrm{CE}}(\mathrm{D}) = -\frac{1}{\mathrm{N}}\left(\sum\nolimits_{\mathrm{pos}}\log\mathrm{f}(\mathrm{x_i}) + \sum\nolimits_{\mathrm{neg}}\log(1-\mathrm{f}(\mathrm{x_i})\right) \qquad (2)$$

where WCE is the weighted cross-entropy, pos represents positive examples, and nos represents negative examples. In cases of class imbalance, a model trained with the standard cross-entropy loss is prone to bias towards the majority class. To counteract this bias, one can employ a weighted version of the cross-entropy loss, which balances the loss contributions from each class. The weighted cross-entropy loss function is defined as

$$\mathcal{L}_{\mathrm{WCE}}(\mathrm{x}) = -\omega_{\mathrm{p}}\mathrm{y}\log\mathrm{f}(\mathrm{x}) - \omega_{\mathrm{n}}(1-\mathrm{y})\log(1-\mathrm{f}(\mathrm{x})) \qquad (3)$$

Here, wp and wn are the weights assigned to the positive and negative classes, respectively. These weights are typically inversely proportional to the class frequencies, thus ensuring that each class contributes proportionally to the total loss, allowing for a more balanced model that fairly represents both classes. In our work, we have integrated this weighted loss function into the HDLaining process. This strategic approach ensures that our model remains sensitive to the less represented class, which is vital for accurate medical diagnostics where the cost of false negatives can be extremely high.

3.3. Classification using hybrid deep learning framework

In this research work, a hybrid deep learning framework was developed to improve early pulmonary nodule detection and malignancy classification from CT imaging. The proposed architecture integrates a Vision Transformer (ViT) backbone with a Enhanced Granular Neural Network (EGNN) module to effectively capture both global contextual dependencies and local structural relationships within lung CT scans. The Vision Transformer component enables the model to learn cross-slice contextual information and identify complex anatomical patterns present in CT images, while the EGNN module constructs connectivity relationships among candidate nodule regions, allowing the network to model spatial interactions and structural associations within lung tissues.

3.3.1. Self-Supervised Graph Vision Transformer

Transformer has shown great success in various natural language processing tasks. Recently, many researchers attempted to explore the benefits of transformer-based models in computer vision tasks [23]. In this work, we introduce Self-supervised vIsion Transformer(SiT) adapted from Vision Transformer (ViT) with some modifications. Vision Transformer receives as input a sequenceof patches obtained by tokenizing the input image x ∈RH×W ×C into n flattened tow-dimensional patches of size p1 ×p2 ×C pixels, where H, W , and C are the height, width, and number of channels of the input image, (p1 × p2) is the patch size, and n is the number of patches, i.e. n = H p1 × Wp2 . Each patch is then projected with a linear layer to D hidden dimensions. The whole operation can also be implemented simply by a convolutional layer with kernel size p1 × p2, number of input and output channels C and D respectively. In order to retain the relative spatial relation between the patches, fixed or learnable position embeddings are added to the patch embeddings as an input to the Transformer encoder.

The Transformer encoder consists of L consecutive Multi-head Self-Attention (MSA) and Multi-Layer Perceptron (MLP) blocks. The MSA layer is defined by h self-attention heads where each head outputs a sequence of size n × d. The employed self attention mechanism is based on a trainable triplet of (query, key, value). Each query vector in Q ∈ Rn×d is matched against a set of key vectors K ∈ Rn×d, and the output is then normalised with a softmax function and multiplied by a set of values V ∈ Rn×d. Thus, the output of the self-attention block is the weighted sum of V as shown in Equation 4. The output sequences of each block are then concatenated into n × dh and projected by a linear layer into n × D sequence.

$$\mathrm{Self\ attention}(\mathrm{Q},\mathrm{K},\mathrm{V}) = \mathrm{softmax}\left(\frac{\mathrm{QK}^{\mathrm{T}}}{\sqrt{\mathrm{d}}}\right)\mathrm{V} \qquad (4)$$

To serve the classification task, a trainable vector (i.e class token) is appended to the input sequence of the patch tokens and goes through the Transformer encoder. Finally, a classification head is added to the output of the Transformer encoder corresponding to the class token. The classification head is implemented by a single linear layer that projects the class embeddings to the number of classes. Unlike ViT this work is based on unsupervised learning/pretraining, hence, classification token is not required. Instead we have a contrastive token, beside the data tokens (image patches token) used for image reconstruction, to serve the proposed self-supervised pretraining tasks.

3.3.2. Enhanced Granular Neural Network (EGNN)

The architecture of a granular neural network along with the associated learning scheme is formed by starting with a certain already designed (numeric) neural network and augmenting its structure by considering granular connections spanned over the numeric weights (connections) [29]. And are concerned with an MLP, which is one of the commonly used topologies of neural networks and comes with a wealth of learning schemes.

A. Interval Operations

Before proceed with the detailed considerations on granular neural networks, let us briefly recall some results of interval mathematics. Given the arguments (variables) represented in the form of intervals, say X = [a, b], Y = [c, d], etc., the algebraic operations are defined as follows:

$$Addition: X + Y = [a + c, b + d] \qquad (5)$$

$$Subtraction: X - Y = [a - d, b - c] \qquad (6)$$

$$Addition: X \times Y = [min(ac, ad, bc, bd), max(ac, ad, bc, bd)] \qquad (7)$$

Division (excluding division by an interval containing 0)

$$\frac{X}{Y} = X \times \left(\frac{1}{Y}\right)\ with\ \frac{1}{Y} = \left[\frac{1}{d}, \frac{1}{c}\right] \qquad (8)$$

Furthermore, when it comes to the mapping (function) of intervals, we have the following main results for non decreasing mapping

$$f(X) = f([a, b]) = [f(a), f(b)] \qquad (9)$$

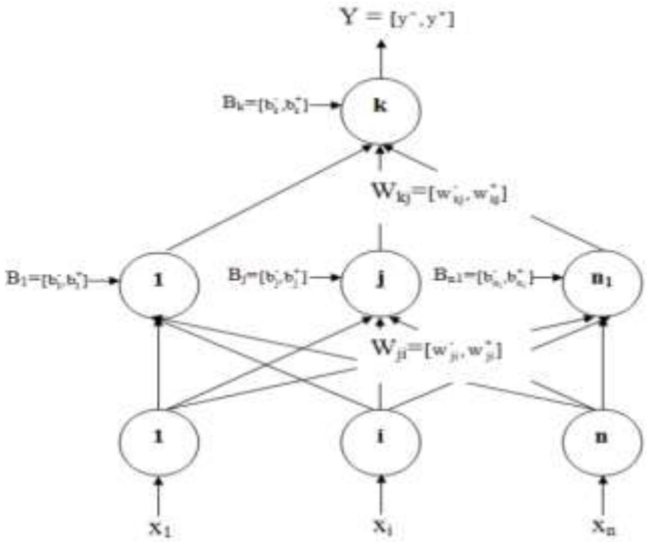
for nonincreasing mapping

$$f(X) = f([a, b]) = [f(b), f(a)] \qquad (10)$$

B. Architectures of Interval-Weight Neural Networks

The granular neural network under consideration, presented in Fig. 3, comprises a single hidden layer consisting of n1 neurons, and a single neuron located in the output layer. Features (inputs) to the network are organized in a vector form x = [x1, x2,. . ., xn]T. The weight (connection) connecting the ith neuron in the input layer to the jth neuron in the hidden layer comes in the form of an interval and is denoted by Wji, Wji =[w− ji,w+ji]. The weights between the hidden layer and

Fig. 3. Architecture of a granular network.



the output layer are also interval valued. Each neuron comes with an interval bias. In virtue of the interval character of the connections used in the network, for any numeric input, the result of processing becomes an interval, say Y = [y−, y+]. As indicated, in design, proceed with the already constructed MLP (e.g., realized in terms of Levenberg–Marquardt backpropagation learning method). The choice of the size of the hidden layer is decided upon during this design phase. The activation functions of the neurons in the hidden layer and the output layer are denoted by f1 and f2, respectively. The detailed formulas are outlined as follows: hidden layer,

$O_j = f_1(z_j), z_j = \sum_{i=1}^{n} w_{ji}x_i + b_j, \qquad j = 1,2,\ldots\ldots, n_1 \qquad (11)$

$f_1(z) = \frac{2}{1+e^{-2\times z}} - 1 \qquad (12)$

output layer

$y = f_2(z), z = \sum_{j=1}^{n_1} w_j o_j + b_j, \qquad f_2(z) = z \qquad (13)$

When it comes to the interval-valued neural network, the aforementioned formulas are generalized in the following way hidden layer:

$o_j = [o_j^-, o_j^+] = [f_1(z_j^-), f_1(z_j^+)] \qquad (14)$

$z_j^- = \sum_{i=1}^{n}(w_{ji}^- x_i + b_j), \qquad (15)$

$z_j^+ = \sum_{i=1}^{n}(w_{ji}^+ x_i + b_j), \qquad j = 1,2,\ldots\ldots, n1 \qquad (16)$

output layer

$Y = [y^-, y^+] = [f_2(z^-), f_2(z^+)] \qquad (17)$

$z^- = \sum_{i=1}^{n}(min(w_j^- o_j^-, w_j^+ o_j^-, w_j^- o_j^+, w_j^+ o_j^+) + b_j) \qquad (18)$

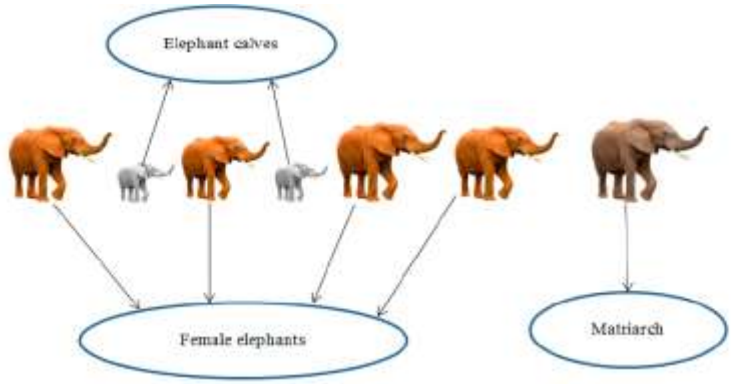
$z^+ = \sum_{j=1}^{n1}(max(w_j^- o_j^-, w_j^+ o_j^-, w_j^- o_j^+, w_j^+ o_j^+) + b_j) \qquad (19)$

As the outputs of the granular neural network are intervals while the targets coming from the experimental data are numeric, we have to define a suitable performance index (objective function), whose optimization (maximization or minimization) is realized through a suitable allocation (distribution) of information granularity. However, the large dataset means the network couldn't trained properly. So this research work introduced a hybrid model for solving the problem. In this work the elephant herding optimization based model is introduced for optimizing the parameter of the network for improving the performance of the classifier.

- **Elephant Herding Optimization (EHO)**

Elephants, as social creatures, live in social structures of females and calves. An elephant clan is headed by a matriarch and composed of a number of elephants [24]. Female members like to live with family members, while the male members tend to live elsewhere. They will gradually become independent of their families until they leave their families completely. The population of all elephants is shown in Figure 4. The following assumptions are considered in EHO.

Figure 4. Population of elephants.



(1) Some clans with fixed numbers of elephants comprise the elephant population
(2) A fixed number of male elephants will leave their family group and live solitarily far away from the main elephant group in each generation.
(3) A matriarch leads the elephants in each clan.

3.5.1. Clan-updating Operator

According to the natural habits of elephants, a matriarch leads the elephants in each clan. Therefore, the new position of each elephant ci is influenced by matriarch ci. The elephant j in clan ci can be calculated by Equation (20).

$$x_{new,ci,j} = x_{ci,j} + a \times (x_{best,ci} - x_{ci,j}) \times r \qquad (20)$$

where $x_{new,ci,j}$ and $x_{ci,j}$ present the new position and old position for elephant j in clan ci, respectively. $x_{best,ci}$ is matriarch ci which represents the best elephant in the clan. $a \in [0,1]$ indicates a scale factor, $r \in [0,1]$. The best elephant can be calculated by Equation (21) for each clan.

$$x_{new,ci,j} = \beta \times x_{centre,ci} \qquad (21)$$

where $\beta \in [0,1]$ represents a factor which determines the influence of the $x_{centre,ci}$ on $x_{new,ci,j}$. $x_{new,ci,j}$ is the new individual. $x_{centre,ci}$ is the center individual of clan ci. It can be calculated by Equation (22) for the d-th dimension.

$$x_{centre,ci,d} = \frac{1}{n_{ci}} \times \sum_{j=1}^{n_{ci}} x_{ci,j,d} \qquad (22)$$

where $1 \leq d \leq D$ and nci indicate the number of elephants in clan ci. $x_{ci,j,d}$ represents the d-th dimension of elephant individual $x_{ci,j}$. $x_{centre,ci}$ is the center of clan ci and it can be updated by Equation (22).

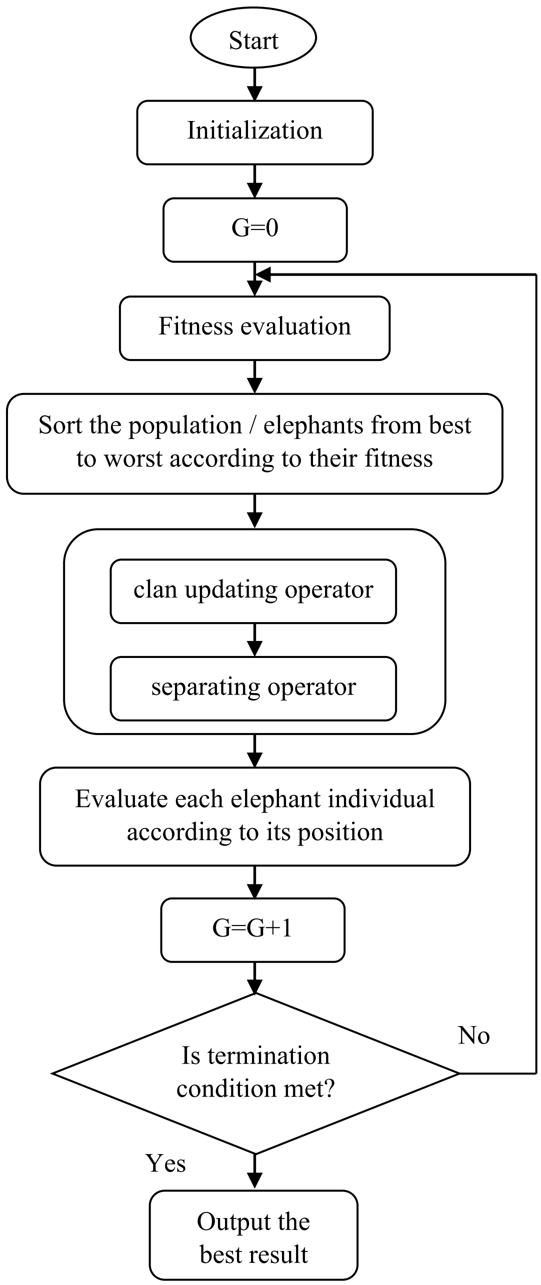
3.5.2. Separating Operator

The separating process whereby male elephants leave their family group can be modeled into separating operator when solving optimization problems. The separating operator is implemented by the elephant individual with the worst fitness in each generation, as shown in Equation (23).

$$x_{worst,ci} = x_{min} + (x_{max} - x_{min} + 1) \times rand \qquad (23)$$

where x_{max} represents the upper bound of the individual and x_{min} indicates lower bound of the individual. $x_{worst,ci}$ indicates the worst individual in clan ci. Rand [0, 1] is a stochastic distribution between 0 and 1. The figure 5. shows the process flow of the elephant herding optimization algorithm.

Figure 5. Flow chart of the elephant herding optimization algorithm



4. Results and Discussion

Then the performance metrics like sensitivity, specificity, accuracy, and precision factor determined using the following equations:

Precision is determined as the ratio of correctly identified positive observations to all of the expected positive observations.

$$\text{Precision} = TP/(TP+FP) \qquad (24)$$

Sensitivity is defined the ratio of correctly identified positive observations to the over-all observations.

$$\text{Recall} = TP/(TP+FN) \qquad (25)$$

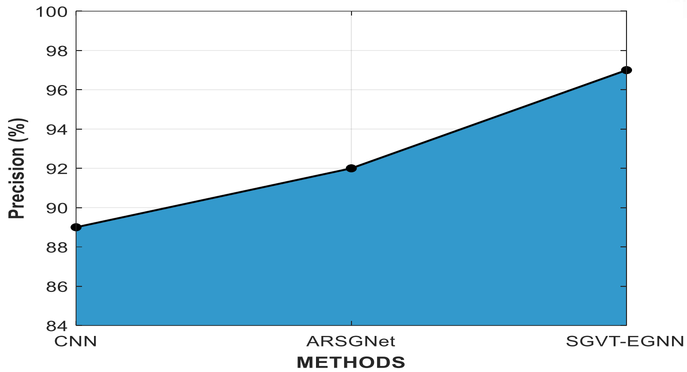
F1 score is determined as the weighted average of Precision as well as Recall. As a result, it takes false positives and false negatives.

$$\text{F1 Score} = 2*(\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision}) \qquad (26)$$

Accuracy is calculated in terms of positives and negatives as follows:

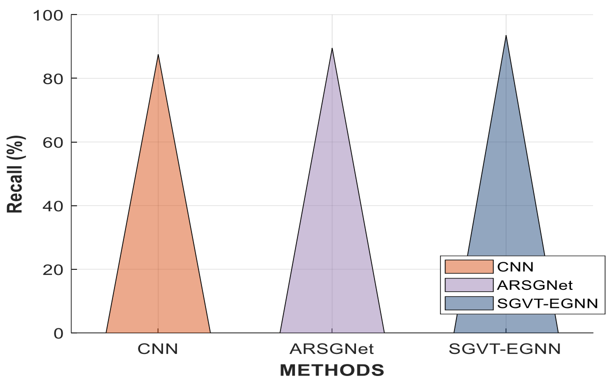
$$\text{Accuracy} = (TP+FP)/(TP+TN+FP+FN) \qquad (27)$$

Figure 6. Precision Comparison of the Proposed and existing method



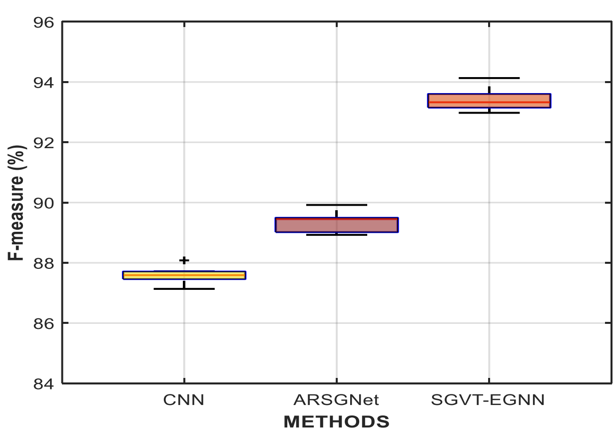
It is noted that, when compared with other existing approaches, best precision is obtained by proposed SGVT-EGNN algorithm based deep learning method. And this preprocessing method increases the overall success rate of the lung disease detection system. From the figure 6, the precision rate of proposed SGVT-EGNN method is 96.7% and ARSGNet method metric is 91.57% and CNN method metric is 88.45%.

Figure 7. Recall Comparison of the Proposed and existing methods



In the Figure.7.shows the recall comparison of the proposed SGVT-EGNN is better when compared with the existing methods. The proposed SGVT-EGNN higher recall results of 93.57%, while the ARSGNet method metric is 89.54% and CNN method metric is 87.68%. It can be seen that the recall increases rapidly at the beginning of training and stabilizes the output with reducing the distance between the points using the proposed SGVT-EGNN based data classification.

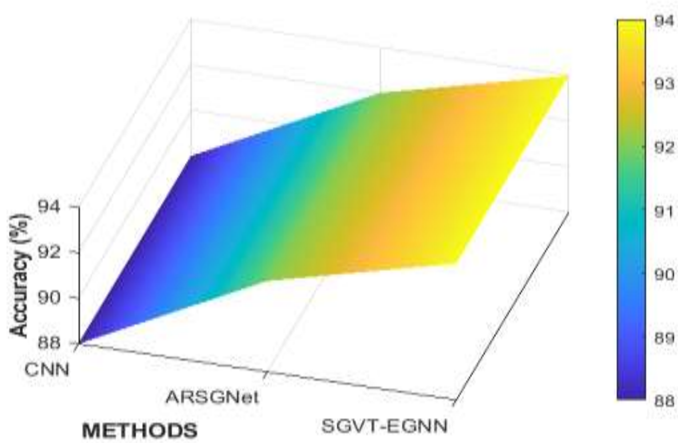
Figure 8. F-measure Comparison of the proposed and existing methods



In the Figure.8.shows the F-measure comparison of the proposed SGVT-EGNN is better when compared with the existing methods. Here the proposed hybrid deep learning process is carried out for the increasing the accuracy of the classifier to avoid the over burden in the classifier stage. It is noted that the proposed SGVT-EGNN method has high f-measure results than the existing methods.

The performance of the proposed SGVT-EGNN method is increases with the help of preprocessing, data imbalance as well classification. Each module has provides the significant improvement in the automated disease detection process. The Fig.9.shows the accuracy comparison of the proposed SGVT-EGNN is better when compared with the existing methods. The proposed SGVT-EGNN produces higher accuracy results of 93.57% where as ARSGNet method metric is 91.81% and CNN method metric is 88.24%.

Figure 9. Accuracy Comparison of the Proposed and exisiting method



5. Conclusion

This research work suggests a Self-Supervised Graph Vision Transformer Network (SS-GVTNet) in detecting and classifying pulmonary nodules and malignancies through CT scanning. The suggested framework combines self-supervised representation learning with the graph-based spatial modeling and transformer-based global feature extraction. By combining transformer-based attention mechanisms with graph-based spatial reasoning, the proposed model enhances feature representation and improves the accuracy of pulmonary nodule detection. The hybrid architecture effectively supports both local feature extraction and global contextual understanding, which are essential for distinguishing benign and malignant nodules in complex medical imaging environments. Experimental evaluation demonstrates that the integrated framework provides improved performance compared with conventional deep learning approaches in terms of detection accuracy, robustness, and computational efficiency. Overall, the proposed method offers a promising computer-aided diagnostic solution that can assist clinicians in the early identification of pulmonary nodules and support timely diagnosis of lung diseases. Future work may focus on incorporating larger multi-institutional datasets, improving model interpretability, and integrating the framework into clinical decision-support systems for real-time medical applications.

References

1. Zhang, G., Jiang, S., Yang, Z., Gong, L., Ma, X., Zhou, Z., ... & Liu, Q. (2018). Automatic nodule detection for lung cancer in CT images: A review. Computers in biology and medicine, 103, 287-300.
2. Kim, T. J., Kim, C. H., Lee, H. Y., Chung, M. J., Shin, S. H., Lee, K. J., & Lee, K. S. (2020). Management of incidental pulmonary nodules: current strategies and future perspectives. Expert review of respiratory medicine, 14(2), 173-194.
3. Tan, Y. (2025, July). Artificial Intelligence in CT-Based Diagnosis of Small Pulmonary Nodules: Current Applications and Future Perspectives. In 2025 10th International Conference on Image, Vision and Computing (ICIVC) (pp. 354-363). IEEE.
4. Hamad, R. H. (2025). A systematic review of artificial intelligence's function in the diagnosis of lung cancer (2018–2024). Mesopotamian Journal of Artificial Intelligence in Healthcare, 2025, 12-25.
5. Liz-López, H., de Sojo-Hernández, Á. A., D'Antonio-Maceiras, S., Díaz-Martínez, M. A., & Camacho, D. (2025). Deep learning innovations in the detection of lung cancer: Advances, trends, and open challenges. Cognitive Computation, 17(2), 67.
6. Megat Ramli, P. N., Aizuddin, A. N., Ahmad, N., Abdul Hamid, Z., & Ismail, K. I. (2025). A systematic review: the role of artificial intelligence in lung cancer screening in detecting lung nodules on chest X-rays. Diagnostics, 15(3), 246.

7. Abdullahi, K., Ramakrishnan, K., & Ali, A. B. (2025). Deep learning techniques for lung cancer diagnosis with computed tomography imaging: A systematic review for detection, segmentation, and classification. Information, 16(6), 451.
8. Kashyap, M., Wang, X., Panjwani, N., Hasan, M., Zhang, Q., Huang, C., ... & Gensheimer, M. F. (2025). Automated deep learning–based detection and segmentation of lung tumors at ct imaging. Radiology, 314(1), e233029.
9. Sandhya, L., & Khan, M. S. (2025). for Pulmonary Nodule Extraction in Lung Cancer Diagnosis. Responsible Artificial Intelligence: Proceedings of ICRAI 2024, 67.
10. Rawat, P., & Kaushik, P. (2025). Global trends and challenges in early lung cancer detection: A bibliometric and systematic approach. Innovations in Computing, 623-628.
11. Safta, W., & Shaffie, A. (2024). Advancing pulmonary nodule diagnosis by integrating Engineered and Deep features extracted from CT scans. Algorithms, 17(4), 161.
12. Wang, G., Lin, Y., Zheng, L., Liang, Y., Zhao, L., Wen, Y., ... & Zhang, L. (2020). A new method for accurately localizing and resecting pulmonary nodules. Journal of Thoracic Disease, 12(9), 4973.
13. Nasrullah, N., Sang, J., Alam, M. S., Mateen, M., Cai, B., & Hu, H. (2019). Automated lung nodule detection and classification using deep learning combined with multiple strategies. Sensors, 19(17), 3722.
14. Shariaty, F., & Mousavi, M. (2019). Application of CAD systems for the automatic detection of lung nodules. Informatics in Medicine Unlocked, 15, 100173.
15. Sori, W. J., Feng, J., & Liu, S. (2019). Multi-path convolutional neural network for lung cancer detection. Multidimensional Systems and Signal Processing, 30(4), 1749-1768.
16. Sang, J., Alam, M. S., & Xiang, H. (2019, May). Automated detection and classification for early stage lung cancer on CT images using deep learning. In Pattern recognition and tracking XXX (Vol. 10995, pp. 200-207). SPIE.
17. Zhang, Q., & Kong, X. (2020). Design of automatic lung nodule detection system based on multi-scene deep learning framework. Ieee Access, 8, 90380-90389.
18. UrRehman, Z., Qiang, Y., Wang, L., Shi, Y., Yang, Q., Khattak, S. U., ... & Zhao, J. (2024). Effective lung nodule detection using deep CNN with dual attention mechanisms. Scientific Reports, 14(1), 3934.
19. Srivastava, D., Srivastava, S. K., Khan, S. B., Singh, H. R., Maakar, S. K., Agarwal, A. K., ... & Albalawi, E. (2023). Early detection of lung nodules using a revolutionized deep learning model. Diagnostics, 13(22), 3485.
20. Mothkur, R., & Veerappa, B. N. (2023). Classification of lung cancer using lightweight deep neural networks. Procedia Computer Science, 218, 1869-1877.
21. Cao, H., Liu, H., Song, E., Ma, G., Xu, X., Jin, R., ... & Hung, C. C. (2020). A two-stage convolutional neural networks for lung nodule detection. IEEE journal of biomedical and health informatics, 24(7), 2006-2015.
22. Oumlaz, M., Oumlaz, Y., Oukaira, A., Benelhaouare, A. Z., & Lakhssassi, A. (2024). Advancing pulmonary nodule detection with arsgnet: EfficientNet and transformer synergy. Electronics, 13(22), 4369.
23. Rong, Y., Bian, Y., Xu, T., Xie, W., Wei, Y., Huang, W., & Huang, J. (2020). Self-supervised graph transformer on large-scale molecular data. Advances in neural information processing systems, 33, 12559-12571.
24. Elhosseini, M. A., El Sehiemy, R. A., Rashwan, Y. I., & Gao, X. Z. (2019). On the performance improvement of elephant herding optimization algorithm. Knowledge-Based Systems, 166, 58-70.