

mU-Net-SAM-ENet: Efficient and Interactive Assistance in Lung Nodule Segmentation

S. Nandini,^{1,4} M. Nagabushanam,¹ Gavisiddappa,² G. S. Nandeesh,³ and M. P. Sundaresha⁴

¹M S Ramaiah Institute of Technology, Bangalore, India

²Channabasaveswara Institute of Technology, Gubbi, India

³Navkis College of Engineering, Hassan, India

⁴Kalpataru Institute of Technology, Tiptur, India

(Received 1 April 2024; Revised 18 May 2025; Accepted 10 June 2025; Published online 8 July 2025)

Abstract: In order to enhance patient recovery, early diagnosis of lung nodules is crucial. Lung cancer continues to be a major worldwide health problem. For lung nodule segmentation, a mU-Net-SAM-ENet is suggested in this work. This work uses a Bayesian threshold-based Taylor series technique for image preprocessing. This method significantly improves contrast and minimizes noise in lung scans, which raises the caliber of the input data. Next, a modified U-Net (mU-Net) architecture, specifically designed for medical picture segmentation, is used in this research. This neural network accurately delineates nodule borders by rapidly capturing complex nodular boundaries. Classification is a crucial stage in the investigation of lung nodules after segmentation. SimAM-EfficientNet, an attention-based deep neural network renowned for its effectiveness and precision in image classification tasks, is used in this method. This research uses the Wildebeest Herd Optimization Algorithm for hyperparameter adjustment to enhance the model's functionality. The experimental study's findings demonstrate that the suggested model, employing the LIDR-IDRI datasets from the Image Database Resource Initiative and the Lung Image Database Consortium, attained accuracy values in both segmentation and classification of 99.76% and 98.2%, respectively.

Keywords: Bayesian threshold-based Taylor series; deep learning; lung segmentation; modified U-Net; SimAM-EfficientNet; U-Net; WHOA

I. INTRODUCTION

One of the most fatal ailments is cancer that strikes people nowadays and poses a threat to their lives. The World Health Organization predicts that by 2020, cancer will be the biggest cause of death globally [1]. Lung cancer is thought to be responsible for the 1.80 million cancer-related fatalities worldwide. Considering projections, cancer would be responsible for 60% of mortality by 2035 [2]. Lung tumors are made up of large numbers of lung cells that have undergone an uncontrolled change. The prevalence of lung cancer has increased globally due to a number of factors. Inhaling hazardous or dangerous substances, which affects a significant number of the senior population in society, is one of the reasons. Smoking cigarettes increases the risk of lung cancer in men by 90% and in women by 70% to 80% [3]. Even those who have never smoked are at danger of developing lung cancer. The most typical lung cancer subtypes are adenocarcinomas and squamous cell carcinomas. Additionally, there are the histological subtypes of small- and large-cell carcinoma. Treatment is more challenging, though, because lung cancers, both big and small cell, can arise in any part of the lung they have a tendency to spread rapidly [4].

When aberrant lung cells multiply uncontrollably and develop into a tumor, squamous cell carcinoma results. One organ where cancer cells can spread (metastasize) is the lymphatic vessels,

which are found around the lungs, liver, adrenal glands, bones, and brain. If left undetected, the condition may develop into squamous cell carcinoma [5]. The incidence and death rates of lung cancer are among the highest of any major cancers in the world as a result. Early identification and therapy of lung cancer are crucial to identification of suspicious lung nodules [6]. The physical characteristics of the malignant cells include an unusually shaped and enlarged nucleus [7]. It is common practice to use the morphological features of computer tomography (CT) scans to differentiate between healthy and unhealthy nodules. Another approach is tried by training the classifiers to detect whether the tissue is cancerous or not. There are other methods that may be applied, including single classifiers and classifiers that combine multiple characteristics [8]. Identifying potential malignant lung nodules is essential for obtaining a lung cancer diagnosis. Lung cancer can be distinguished from benign nodules by contrasting their features [9]. For an appropriate nodule evaluation, careful examination of the morphologic features is necessary because There is significant overlap between the traits of nodules that are both benign and malignant. Additionally, morphological analysis is necessary for early diagnosis [10].

The identification, segmentation, and categorization of benign and malignant nodules in the lungs are the key research areas for the application of imaging techniques based on deep learning. To enhance the functionality of deep learning models, research mostly focuses on creating novel designs for networks and loss functions. Recent reviews of deep learning approaches have been published by a number of research organizations [11–13]. However, deep

Corresponding author: S. Nandini (e-mail: nandinipinky6@gmail.com).

learning techniques have advanced quickly, and a number of fresh ways and uses appear every year. The study's remaining sections are structured using shadows: The relevant works are summarized in Section II, the suggested model is briefly explained in Section III, the findings and validation analysis are shown in Section IV, and the summary and conclusion are given in Section V.

II. RELATED WORKS

In their work, Li, X., *et al.* [14] imaginatively proposed a Bidirectional Pyramid component and a Cross-Transformer module to successfully circumvent the aforementioned two shortcomings, by merging its self-awareness on the inside and outside. The latter minimizes feature inconsistencies at various stages of the model by incorporating bidirectional entry points into the characteristic pyramid that are used to encourage feature information exchange between the shallow and deeper layers of the model. A comprehensive lung nodule analysis significantly improves overall performances such as Dice Similarity Coefficient and sensitivity. The model put forward in this research has exceptional lung nodule segmentation performance, and it can assess the size, form, and other characteristics of lung nodules, which is very useful for clinicians in the early identification of lung nodules and has tremendous clinical importance.

In their paper, Annavarapu, C.S.R. *et al.* [15] created a resource-effective end-to-end deep learning method for segmenting lung nodules. Between an encoder and a decoder, a bidirectional feature network is used. To increase the effectiveness of the segmentation, it also makes use of the Mish function of activation and the class strengths of masks. Using the 1186 lung nodules in the publicly available LUNA-16 dataset, the suggested model was thoroughly tested and trained. The network training parameter was each training sample's weighted binary cross-entropy loss to increase the probability that each voxel in the mask belongs to the correct class. Additionally, the suggested model's robustness was assessed using data from the QIN Lung CT project. The evaluation compares the performance of competing deep learning models and provides the U-Net as better than other deep learning models.

The study of Mothkur, R., and Veerappa, B.N. [16] set out to evaluate various deep learning architectures for image processing that are memory-efficient and portable. With an accuracy of 85.21 percent, the suggested compact deep neural network outperforms by making an appropriate trade-off between specificity and sensitivity. The proposed study employs binary classification networks to distinguish between lung cancer in its early stages and other types of cancer in patient CT pictures. These networks include 2D SqueezeNet, 2D MobileNet, and plain old 2D CNN.

Tyagi, S., *et al.* [17] devised a technique for segmenting lung tumors that combines the convolutional neural network and the vision transformer. Convolution blocks are employed in the earliest layers of the encoder and the decoder's last layers, which capture features that communicate the most crucial information. This network is designed as an encoder-decoder structure. Transformer blocks with self-attention mechanisms are used in the deeper layers to produce more precise global feature maps. For network optimization, they make use of a recently proposed unified loss equation that includes dice-based and cross-entropy losses. After training it using an NSCLC-Radiomics dataset that is accessible to the public, they evaluated the generalizability of their network using a dataset they obtained from a neighboring hospital. The results are based on the average Dice coefficient and average hausdorff distance for each method, respectively.

Techniques like enhanced backpropagation of feed-forward neural network (EBFNN), convolution neural network (CNN), and random forest (RF) were used by Wahengbam and Sriram [18] in their investigation. When these algorithms' accuracy was evaluated, it was discovered that EBFNN outperformed other methods. The four essential processes in the execution of the suggested model are preprocessing, segmentation, feature elimination, and classification. Utilizing a common set of images gathered by the Lung Image Database Consortium (LIDC), the implementation step's results are reviewed from a number of angles. The results show that, compared to RF accuracy levels of 89% and CNN accuracy levels of 91%, the recommended EBFNN provides accuracy levels of about 93%.

Using both whale optimization and bacterial foraging optimization, Alameen, A.'s study [19] aimed to create a fused evolutionary technique that enhances feature extraction. For the classification process, a dual-graph artificial neural network (CNN) was employed. The study evaluate the fused model's performance using several metrics and methods. Initially, the input CT image is used to determine whether the lesion is malignant or not. A number of methods, including the Histogram-oriented gradient features, gray-level co-occurrence matrix, and Gray-level dependency matrix, are applied to the pre-processed segmented picture to extract its geometrical, statistical, structural, and texture features. The best features should be chosen using a brand-new fusion method called Whale-Bacterial Foraging Optimization. Lung cancer classification has been carried out using dual-graph convolutional neural networks. Algorithms for classification and optimization have been compared in study.

III. PROPOSED METHODOLOGY

Figure 1 represents the pipeline of the proposed work.

A. DATASET DESCRIPTION

The LIDC-IDRI dataset was used to create the suggested model [20], which has 1012 variable-slice-thickness LDCT scans. Diameter-specific nodules were verified to confirm whether they were actual nodules. The collection also includes a numerical description of a number of nodular traits. Specifically, the transparency of the nodules is indicated by the nodule texture, where 1 denotes an entirely solid nodule and five entirely non-solid nodules. 2 284 nodules were evaluated from the 888 images used for the LUNA16 challenge 18 (due to irregular annotations, inadequate scan reconstruction, or high slice thickness, several samples were discarded). Of those, 1 593, 1 190, and 790 had a level 2, level 3, and level 4 of agreement. Nodules are classified as sub-solid in our testing if its average consistency is lower than 5, solid in our tests if it is more than 5, and non-solid in all other cases. There are 1695 solid nodules in the dataset, with a common agreement level of 2, 300 sub-solid nodules, and 135 non-solid nodules [21].

The average center of masses of the specialized annotations served as the patching point for all nodules, which were subsequently isotopically scaled to $64 \times 64 \times 64$ voxels. Volume picture intensity was linearly transformed from [400, 1000] to [0, 1] Hounsfield Units. The optimizer Wildebeest Herd Optimization Algorithm (WHOA) was employed (learning rate 0.001), and an 8-sample batch size was employed to train the network. To accommodate for inter-observer variability, all agreement levels were taken into account; nonetheless, the algorithm was trained by matching the same nodule with a number of reliable ground truths.

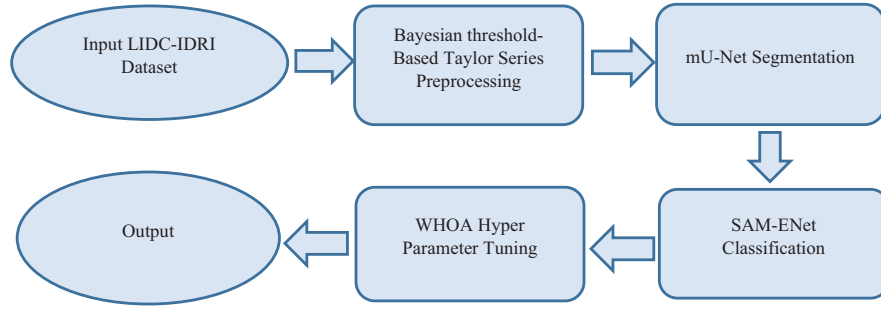


Fig. 1. Proposed workflow.

Additionally, this paper used 20% of the training for validation and evaluated proposed model using stratified five-fold cross-validation using scan-level partitions. Using 100 search steps and a random search20, all hyper-parameters were located. At each step, $\{\lambda_1, \gamma, p\} \sim U([0,1])$, where U is a uniform distribution. The initial train-test split's validation set was subjected to optimization.

B. PREPROCESSING

The first step in denoising lung pictures is to enhance their quality [22]. The Bayesian threshold-based Taylor series strategy is suggested in this study as a method for turning noisy pixels into pixels of higher quality. Since the Taylor series offers high-resolution wavelet subbands, this suggested method is based on it. An infinite number of terms resulting from derivatives are defined by the Taylor series with pixel-based and changeable values. A 3D frame is required for the Taylor series as equation (1).

$$T(X) = (a) + (X - a)\hat{T}Df(a) + 1/2!(X - a)\hat{T}\{D^2f(a)\}(X - a) + \dots \quad (1)$$

Since an image's intensities are not uniform throughout, images typically have a variety of pixel values. For denoising, the variations in pixel values therefore require greater consideration. The noise in subbands is calculated from the defined Taylor series and removed if it exceeds the predicted Bayesian threshold. The following equation (2) is used to calculate the threshold for all subbands:

$$T_B = \frac{\sigma_N^2}{\sigma_S} \quad (2)$$

where σ_N indicates the assessment of the noise, and T_B the threshold was reached for all bands.

When $\sigma_S \neq 0$, It is written as equation (3).

$$\sigma_s = \sqrt{\max((\sigma_y^2 - \sigma_y^2), 0)} \quad (3)$$

where σ_y is defined as equation (4).

$$\sigma_y = \frac{1}{N} \left(\sum SB_i \right) \quad (4)$$

The subbands SB_i are $SB_i = \{\{LLL, LLH, LHL, LHH, HLL, HLLH, HHL, HHH\}\}$. This is regarded as the overall number of subbands, and σ_N is influenced by equation (5).

$$\sigma_N = \left[\frac{\text{median}(SB)}{0.6745} \right] \quad (5)$$

After calculating the T_B following an arrangement of the acquired Bayesian threshold values, an expression is created by applying the curve fitting technique by equation (6):

$$\gamma = \frac{p_1\beta^2 + p_2\beta + p_3}{\beta + q} \quad (6)$$

In Equation (6), β provides a definition of the standard deviation of the noise and the amounts of other variables such as $p_1 = 0.9592$, $p_2 = 3.648$, $p_3 = -0.138$, and $q = 0.1245$. The mathematical formulation of the BM3D filtering technique is then used as equation (7):

$$\hat{Y}_{SB} = T_{3D}^{-1} \left(\alpha \left(T_{3D}(Z_{SB}), T_B \gamma \sqrt{2 \log \log(N^2)} \right) \right) \quad (7)$$

In Equation (7), the T3D is defined as single 3D transform. Each subband is reconfigured and achieves denoised frames from this final formulation. The process of intensity normalization, which follows, is carried out with a range between 0 and 1. Intensity normalization uses the Z score normalization function, whose performance is superior to that of the decimal and min-max normalization methods. The following is how this normalization function is calculated:

$$\gamma_i = \frac{x_i - \mu}{\delta} \quad (8)$$

Where γ_i represents the voxel intensity at location k, variance as δ and intensity's mean value as μ . When preparing an image for subsequent processing, the rational range (0–1) was employed for intensity normalization using the Z score normalization tool. For operations involving training and testing, this phase is required for all photos.

C. mU-NET SEGMENTATION

To avoid repeating low-resolution feature map data, the network that is suggested uses a residual path. After pooling in the potential network, the remaining route is positioned to the right, as seen in Figs. 2(b)–(d), in contrast to the prior study from [23]. After pooling, the remaining layers are placed to right side position and it is shown in Figs. 2(b)–(d), which is contrasted from the previous study [23]. This adaptive filter's performance was evaluated using a predetermined permeation rate as equation (9):

$$\text{Permeation rate} = \begin{cases} -0.5, & \text{if } FM_b(x,y) < 0.01 \\ \frac{\sum_{x,y \in \text{object}} \text{label}(x,y) FM_b(x,y)}{\sum_{x,y \in \text{object}} \text{label}(x,y) FM_b(x,y)}, & \text{otherwise,} \end{cases} \quad (9)$$

where FM_a is the residual path being followed by the normalized feature map in the skip connection. FM_b , where each label

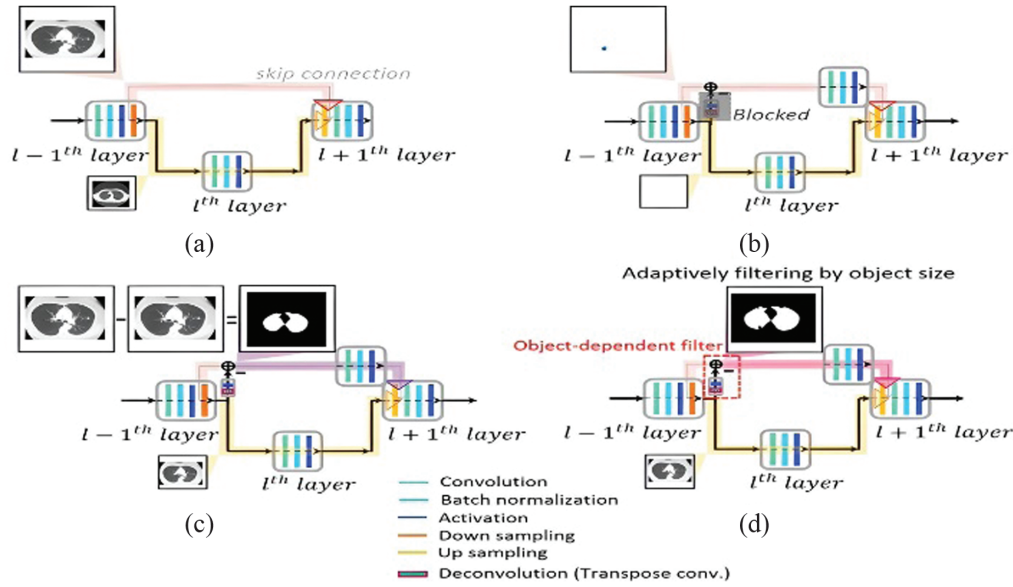


Fig. 2. Diagram showing the proposed mU-Nets in (b) and (d) and the standard U-Net in (a). (a) When using a traditional U-Net, all feature information travels only low-resolution data transmitted to the next level via the skip link. Small objects' spatial information frequently vanishes after pooling owing to resolution loss. (b) For small objects in the mU-Net case, pooling allows for the extraction of higher-level global characteristics without sacrificing resolution. Small object characteristics can enter the connection that skips without being deleted by pooling because the deconvolution path is blocked, preserving the location data of small objects. (c) To prevent the repetition of low-resolution information, feature information for large objects in the mU-Net example is limited to edge information in the skip connection. (d) The suggested network is represented schematically. Depending on the size of the item, deconvolution and activation in the residual path merge leftover path elements into skip connection features.

represents the binary mask of the item (lung or lung tumor) and x, y presents the residual path and a standardized feature map. The range of each normalized feature map is $[0, 1]$. The penetration rate was set to -0.5 when $FM_b(x, y) < 0.01$. This means that the skip connection lacks any useful functionality.

Because pooling destroys spatial information, as illustrated in Fig. 2(b), a few pooling layers may be sufficient to extract the small object's overall features. Less information is wasted due to pooling as seen in Fig. 3, which results in more effective feature extraction for small objects. As illustrated in Fig. 2(b), the residue path is used to preserve tiny object information and subsequent convolution layers to the bypass connections to extract greater-level features. This adjustment allows penetration rates of minor object features to remain high at early stages while preserving information that would be lost following additional poolings, as illustrated in Fig. 3. As the stage grows, information on little objects eventually disappears, as seen in Fig. 3 with penetration rates of -0.5 . Large object features should only contain edge information in the skip connection, as shown in Fig. 2(c), to increase the effectiveness for huge object feature extraction. Since the poor resolution information has already spread to later stages, as seen in Fig. 2(a), there is less need to remove it. The stage in this study is connected to the size of the feature matrices. In other words, features are extracted in the same stage with the same matrix size. At each stage, the size represented by the blue dashed lines is 2828 mm^2 .

Upsampling, carried out through a layer of deconvolution along the residue path, as well as residual operation at the connection between skips, as illustrated in Fig. 2(d), automatically removes information based on object size. c_u^{l+1} that is a signal in the remaining course after the suggested object-dependent upsampling, and c_s^{l+1} that is a signal defined following the shortening by the omission of a convolution layer's skip connection as follows by equations (10) and (11),

$$\begin{aligned} c_{u,v}^{l+1} &= f(w_{u,v}^{l+1} \otimes^T D(c^{l-1}) + b_{u,v}^{l+1}) \in c_u^{l+1}, \\ c_u^{l+1} &\equiv C_u^{l+1}(D(c^{l-1}); \theta_u^{l+1}), \end{aligned} \quad (10)$$

$$\begin{aligned} c_{s,t}^{l+1} &= f(w_{s,t}^{l+1} \otimes (c^{l-1}) + c_u^{l+1}) + b_{s,t}^{l+1} \in c_s^{l+1}, \\ c_s^{l+1} &\equiv C_s^{l+1}(c^{l-1} - c_u^{l+1}); \theta_s^{l+1}, \end{aligned} \quad (11)$$

Where reversible convolution as $\otimes^T$, kernel weighting as w , bias worth as b and sampling set as θ in the skip link layer of residual path. All the weight, bias, kernel and sampling set in equation (11) are the complete set of convolutional layer parameters, where they are different from each other. Subsequently, the prior signal is characterized by equation (12) that is prior to the convolution process at the decoding stage and it is characterized in Fig. 2(d),

$$\begin{aligned} p_{\text{proposed}}^{l+1} &= c_s^{l+1} \oplus u(c^l) \\ &= c_s^{l+1}(c^{l-1} - c_u^{l+1}; \theta_s^{l+1}) \oplus u(C^l(D(c^{l-1}); \theta^l)). \end{aligned} \quad (12)$$

From equation (12), adaptive filtering is used to filter feature data in the residual path $c^{l-1} - c_u^{l+1}$ in contrast to the traditional U-Net in equation (6). Let's assume, for the sake of simplification, that adaptive up-sampling causes characteristics of the enormous object to be interpolated and features from the small object to be annihilated filtered at the early stage. The following are the results of the suggested network. By equations (13) and (14),

$$\begin{aligned} p_{\text{proposed}}^{l+1} &= c_s^{l+1}(c^{l-1} - c_{\text{low}}^{l-1}; \theta_s^{l+1}) \oplus u(C^l(D(c^{l-1}); \theta^l)), \\ c_u^{l+1} &= u(D(c^{l-1})) \equiv c_{\text{low}}^{l-1} \end{aligned} \quad (13)$$

$$p_{\text{proposed}}^{l+1} = c_s^{l+1}(c^{l-1}; \theta_s^{l+1}) \oplus u(C^l(D(c^{l-1}); \theta^l)), \quad (14)$$

where $c_u^{l+1} = 0$.

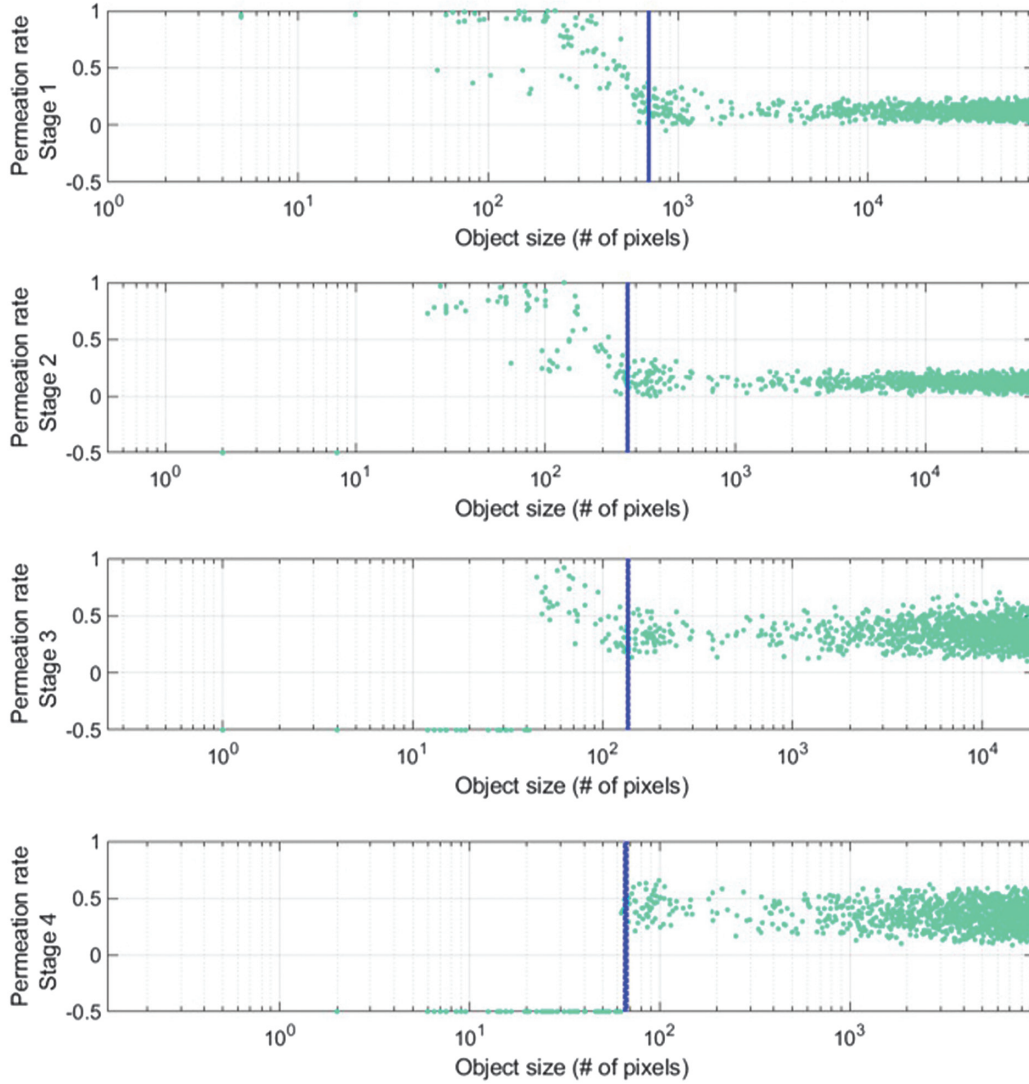


Fig. 3. The mU-Net's permeation rate in relation to phases.

As seen in Fig. 3, edge-like information can be found in feature maps of large objects formed here after the residual pass in the bypassing connection. The suggested network extracts edge data for large objects with a preference that is matched to $c_u^{l+1} = u(D(c^{l-1})) \equiv c_{low}^{l-1}$ in equation (13). Contrarily, small object feature maps extract global features that are matched to the target better and do not have identical resolving losses as a normal U-net architecture (loss from deconvolution and pooling are absent) $c_u^{l+1} = c^{l-1}$ in equation (14). In addition, both $c^{l-1} - c_{low}^{l-1}$ and c^{l-1} in equations (13) and (14) pass the further convolutional layer of c_s^{l+1} for the purpose of extracting higher-level features, skipping the connection. The suggested network additionally makes use of weight decay, dropout, and batch normalization to increase accuracy. The suggested network's loss function is defined as the mean square error (MSE) over the outputs of estimate and desired multi-class segmentation. Figure 4 shows the architecture of mU-Net segmentation [24].

D. SAM-ENet CLASSIFICATION

1. ENet MODEL. A new convolutional neural network called ENet was proposed by Google in May 2019. ResNet

deepens the network primarily to improve accuracy. ENet balances the resolution, depth, and width of the network using compound scaling factors to improve accuracy and performance.

2. SAM ATTENTION MODULE. Various factors influence and complicate the topologies of various convolutional neural networks. SimAM, in comparison, considers the channel as well as the space. Without changing the original network's parameters, three-dimensional attention weights are suggested. It develops a quickly convergent solution while defining a neuroscience-based energy function particularly. SimAM frequently outperforms the most common attention modules for SE and CBAM. SimAM can perform better than these two common categories of attention modules, as illustrated in Fig. 5 [25].

The SimAM module establishes an energy function and searches for significant neurons. It adds regular words and binary labels. Finally, the following formula can be used to get the lowest energy by equation (15):

$$e_i^* = (4(\lambda + \sigma^2)) / ((t - u)^2 + 2\sigma^2 + 2\lambda) \quad (15)$$

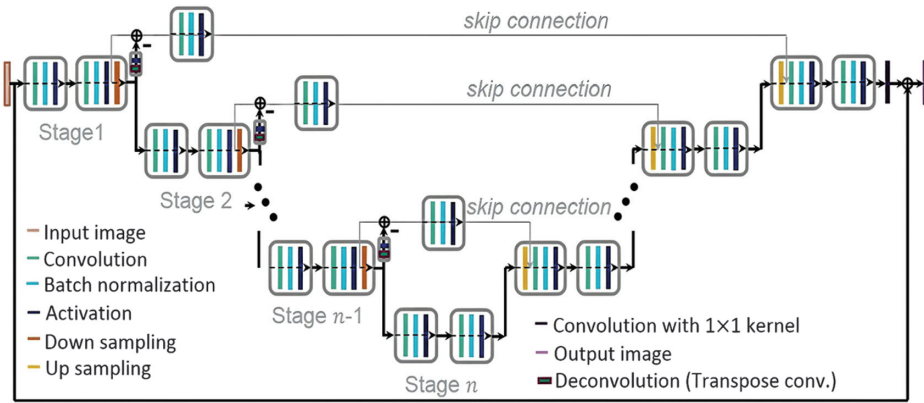


Fig. 4. mU-Net architecture.

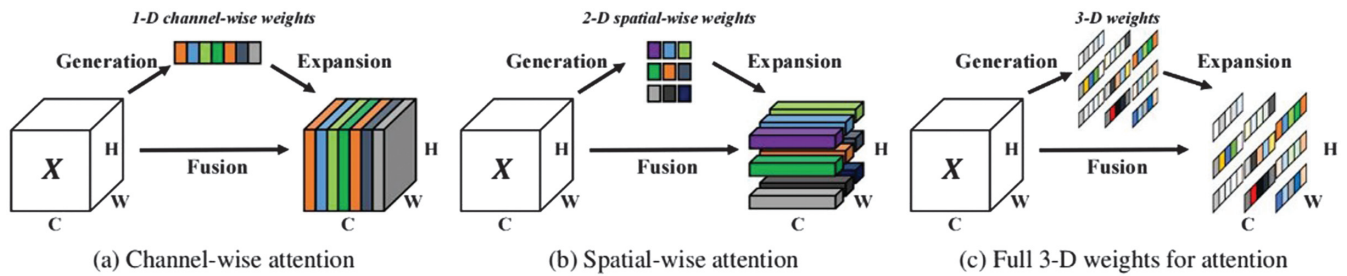


Fig. 5. Three different attention steps are compared.

$$\text{where } u_t = \frac{1}{M-1} \sum_{i=1}^{M-1} x_i, \quad \sigma_t^2 = \frac{1}{M-1} \sum_{i=1}^{M-1} (x_i - u_t)^2 \quad (16)$$

One of them, μ_t is the average neuronal value. σ_t^2 is a measure of all neurons' variance t , the intended neuron. x_i the input feature's channel has additional neurons. λ is a measure of regularization. You can determine how many neurons are present on that channel by $M = H \times W$. In conclusion, energy plays a role in the distinction between neurons and neurons in the peripheral cortex. Each neuron's importance can be determined by $1/e^*$. To fine-tune features, it employs a scaling operator. The SimAM module's entire development process is shown in equation (17):

$$X = X \cdot \text{sigmoid}\left(\frac{1}{E}\right) \quad (17)$$

where E aggregates all e in the spatial and channel dimensions. To restrict the size of the value in E , sigmoid is used.

3. SAM-ENet. There is a 3×3 convolution layer in the first stage. MBConv, the most significant structural component of this network model, is present in the second through eighth stages. One-to-one convolution, pooling, and the final completely connected layers make up the final stage. MBConv is divided into five sections. A 1×1 convolution layer makes up the first component. A depth-wise convolution comes next and then the SE attention module. A 1×1 convolution layer is the fourth component, which reduces the number of dimensions. A dropout layer comes last as a solution to the overfitting issue, in order to boost the channel and spatial weights. Following the initial convolutional layer was the SimAM module. Figure 6 [26] depicts the fundamental structure of this paper,

MBConv. It suitably modifies the remainder of the structure. The SE attention module is part of the original ENet. To make it better, our model incorporates the SimAM attention module.

In this study, SAM-ENet is made up of two layers of convolution, one pooling layer and a single fully connected layer in addition to seven upgraded SAM-MBConv modules. In Fig. 7, different levels are represented by various colors and sizes. First, the pictures with the dimensions of $224 \times 224 \times 3$ are ascended by the 3×3 convolutional layer. The size of the feature-filled images that were acquired is $112 \times 112 \times 32$. The features of the photos are then extracted using SimAM-Conv. When two SimAM-Conv are identical, the connection is terminated and the input is established. After 1×1 The original channel is recovered using point-by-point convolution, and classification is performed on the fully linked layer.

4. WILDEBEEST HERD OPTIMIZATION ALGORITHM. The WHOA [27] has been used in this study to optimize our deep network, and it was chosen primarily because it is one of the more modern types of metaheuristic algorithms. The WHO algorithm was influenced by wildebeests' foraging behavior, and its benchmark function findings, which were based on the research, likewise display superior results. This obliges us to utilize this metaheuristic approach to increase the effectiveness of the suggested CNN. Male sex challenges with rivals are used by wildebeests, a sociable mammal species that travels in search of food sources, in order to entice females for mating.

The WHO algorithm initializes a number of populations (wildebeests) at random as potential possibilities. There are only a few people in the lower classes ($X_{\min}$), and the higher ($X_{\max}$) boundaries, i.e.,

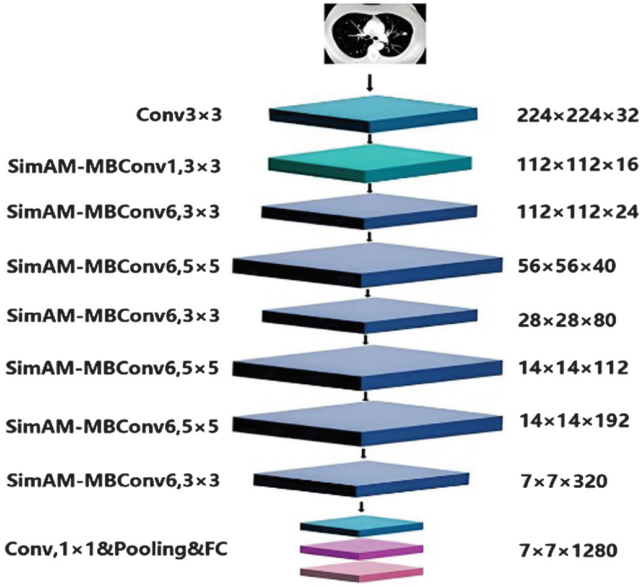


Fig. 6. MBConv modified.

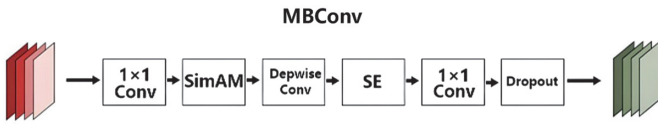


Fig. 7. SAM-ENet architecture.

$$X_i \in [X_{\min}, X_{\max}] \quad (18)$$

where $i = 1, 2, \dots, N$.

The wildebeest then employs the milling process to move about locally. The phase is depicted by continuing to search for the ideal position while taking into account a fixed number (n) is used as the little random movement that varies with the solution spaces, a haphazard phase Z_n has been used by the applicants for position X to often look for the little random phase opportunities. The candidate results were adjusted using a random step size in a customizable length. The local experimentation phase as a result Z_n obtained as shown in the following equation (19):

$$Z_n = X_i + \varepsilon \times \theta \times v \quad (19)$$

where ε represents the factor affecting learning speed, X_i the candidate number is described i , θ signifies a randomly generated, uniformly distributed number between 0 and 1, where v defines a vector of random units.

The wildebeest then adjusts its position to obtain the best possible random location after evaluating a fixed number (n) of unimportant candidates at random, i.e.,

$$X_i = \alpha_1 \times Z_n^* + \beta_1 \times (X_i - Z_n^*) \quad (20)$$

where α_1 and β_1 elucidate the factors that the leaders will use to direct the candidates' local movements.

The previous step is to mimic the swarm instinct of wildebeests. Once the other contenders are situated in an area with an appropriate food source, this is mimicked, i.e.,

$$X_i = \alpha_2 \times X_i + \beta_2 \times X_h \quad (21)$$

where X_h establishes a random candidate, α_2 and β_2 indicate the factors that the leader will use to direct the crew's local movement.

There is additional term in the WHOA to prevent the candidates from traveling to areas with limited food sources. The mathematical model for this word is equation (22):

$$X_i = X_i + \theta \times (X_{\max} - X_{\min}) \times v, \quad (22)$$

where v describes a unit vector of randomness.

The algorithm also includes a phrase for simulating busy areas. When the grass is fertile on a large scale, there is a throng. The name of this concept is "individual pressure." This word allows a challenge to be completed and the strongest contender can destroy the others utilizing equation (23):

$$\begin{aligned} \text{If } (\|X^* - X_i\|) < \eta, \\ (\|X^* - X_i\|) > 1, \quad \text{then: } X_i = X^* + \varepsilon \times \hat{n}, \end{aligned} \quad (23)$$

where η indicates a threshold so that people won't swarm into the places and $\hat{n}$ describes the approximate amount of exploitable components.

The swarm social memory, which is simulated in the last stage to provide better placements, is obtained via the following methods equation (24):

$$X = X^* + 0.1 \times \hat{v} \quad (24)$$

The total output provides an overview of the system "Wildebeest Herd Optimization".

- (1) setting up answers for the Wildebeest Herd. When hunting for the solution space, the source of the fertile grass, the herd follows the leader.
- (2) attempting to capture the most productive source of food for the wildebeest based on the algorithm's conditions for exploitation and exploration.
- (3) As long as the termination conditions have not been met, all stages are repeated.

IV. RESULTS AND DISCUSSION

A. EXPERIMENTAL SETUP

On a desktop with an Intel Core i7-5960X, 32 GB of RAM, and two GTX 1080 graphics cards, experiments were run using Python 3.5 and Keras 2.2.

Table I. Comparison of segmentation models

Models	Accuracy	Precision	Recall	F1-Score	Specificity
Mask R-CNN	87.33	86.95	74.58	83.712	84.23
DenseNet	88.9	81.06	79.21	87.15	88.32
SegNet	93.79	85.47	83.22	86.43	91.54
U-net	95.16	89.11	89.45	90.65	95.86
Proposed model	99.76	99.85	99.68	99.66	99.32

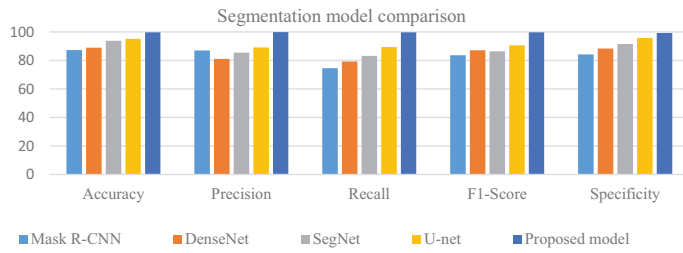


Fig. 8. Analysis of segmentation models.

B. PERFORMANCE METRICS

From Table I, Fig. 8, the mU-net model's multi-scale, multi-modal approach assists in lowering false positives, a crucial component in lung nodule segmentation to assure correct diagnosis and lower the need for pointless treatments, which produces better results for the proposed model than for other models.

Insights from Table II and Fig. 9, the proposed model achieved 98.2% of accuracy, 97% of precision, 97% of recall, 97.8% of F1-score and 97.5% of specificity, that shows better performance than the cited articles of ACO, GWO, and ABC Algorithms.

V. CONCLUSION

Our research on lung nodule segmentation using the LIDC-IDRI dataset has yielded promising results and significantly improved the field of medical image analysis. This paper introduced the Bayesian threshold-based Taylor series technique for picture pre-processing. This paper developed a state-of-the-art technique to enhance lung nodule images while reducing noise. The segmentation method employed in this work, the mU-net design, is famous for being effective at segmenting medical images. These segmented lung nodules were categorized with the assistance of the cutting-edge neural network architecture SAM-ENet. This has a very high degree of classification accuracy for nodules. Our models were carefully calibrated and optimized for the task at hand using the innovative WHOA for hyperparameter tuning in this work. As a result of the experimental study, it was determined that the suggested model had segmentation and classification accuracy values of 99.76% and 98.2%, respectively. Even if the results of our recommended strategy were good, there is still room for advancement. Evaluations should be performed on large and diverse datasets to improve the system's resilience.

ACKNOWLEDGEMENT

The authors acknowledge the National Cancer Institute and Foundation for the National Institute of Health for their critical role in the creation of publicly available LIDC-IDRI and RIDER databases used in this study.

Table II. Models for classification are compared

Models	Accuracy	Precision	Recall	F1-Score	Specificity
AlexNet	86.5	87	86	86	86
GoogleNet	87.5	91	87	89	88
ResNet-50	91.7	92.5	92.5	93.4	95
VGG-16	94	94.2	95	94.6	96.3
Proposed model	98.2	97	97	97.8	97.5

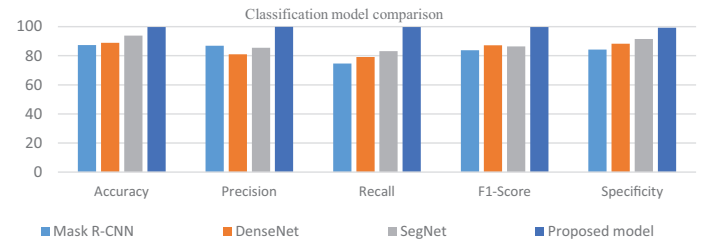


Fig. 9. Comparison of classification models.

DATA AVAILABILITY STATEMENT

The datasets generated and/or analyzed during the current study are available in the Cancer Imaging Archive (TCIA) repository, <https://www.ncbi.nlm.nih.gov/pubmed/23884657>.

AUTHOR CONTRIBUTION

Nandini and Nagabushanam conceived the idea and supervised the findings of this work. Nandeesh, Gavisiddappa and Sundaresha assisted with the language of the manuscript. All authors discussed the results and approved the final manuscript.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no competing interest.

FUNDING

Not Applicable.

ETHICAL APPROVAL AND CONSENT TO PARTICIPATE

This article does not contain any studies with human participants or animals performed by any of the authors.

CONSENT TO PUBLICATION

Not Applicable.

REFERENCES

- [1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA Cancer J. Clin.*, vol. 71, pp. 209–249, 2021.

- [2] M. Araghi, I. Soerjomataram, M. Jenkins, J. Brierley, E. Morris, F. Bray, and M. Arnold, "Global trends in colorectal cancer mortality: projections to the year 2035," *Int. J. Cancer*, vol. 144, pp. 2992–3000, 2019.
- [3] M. Masud, N. Sikder, A.-A. Nahid, A. K. Bairagi, and M. A. AlZain, "A machine learning approach to diagnosing lung and colon cancer using a deep learning-based classification framework," *Sensors*, vol. 21, p. 748, 2021.
- [4] E. A.-R. Hamed, M. A.-M. Salem, N. L. Badr, and M. F. Tolba, "Lung cancer classification model using convolution neural network," In *Proceedings of the 3rd International Conference on Artificial Intelligence and Computer Vision (AICV2023), Marrakesh, Morocco, 5–7 March 2023*. Springer: Berlin/Heidelberg, Germany, 2023, pp. 16–26.
- [5] S. K. Thakur, D. P. Singh, and J. Choudhary, "Lung cancer identification: a review on detection and classification," *Cancer Metastasis Rev.*, vol. 39, pp. 989–998, 2020.
- [6] F. Carrillo-Perez, J. C. Morales, D. Castillo-Secilla, O. Gevaert, I. Rojas, and L. J. Herrera, "Machine-learning-based late fusion on multi-omics and multi-scale data for non-small-cell lung cancer diagnosis," *J. Pers. Med.*, vol. 12, no. 4, p. 601, 2022.
- [7] M. Wei, Y. Du, X. Wu, Q. Su, J. Zhu, L. Zheng, G. Lv, and J. Zhuang, "A benign and malignant breast tumor classification method via efficiently combining texture and morphological features on ultrasound images," *Comput. Math. Methods Med.*, vol. 2020, p. 5894010, 2020.
- [8] S. Zheng, J. Guo, X. Cui, R. N. Veldhuis, M. Oudkerk, and P. M. van Ooijen, "Automatic pulmonary nodule detection in CT scans using convolutional neural networks based on maximum intensity projection," *IEEE Trans. Med. Imaging*, vol. 39, pp. 797–805, 2019.
- [9] P. Afshar, A. Oikonomou, F. Naderkhani, P. N. Tyrrell, K. N. Plataniotis, K. Farahani, and A. Mohammadi, "3D-MCN: a 3D multi-scale capsule network for lung nodule malignancy prediction," *Sci. Rep.*, vol. 10, p. 7948, 2020.
- [10] M. Mandal and S. K. Vipparthi, "An empirical review of deep learning frameworks for change detection: model design, experimental frameworks, challenges and research needs," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, pp. 6101–6122, 2022.
- [11] H. Alireza, M. Cheikh, A. Annika, and V. Jari, "Deep learning for forest inventory and planning: a critical review on the remote sensing approaches so far and prospects for further applications," *For.*, vol. 95, pp. 451–465, 2022.
- [12] F. Highamcatherine and J. Highamdesmond, "Deep learning," *SIAM Rev.*, vol. 32, pp. 860–891, 2019.
- [13] X. Li, A. Jiang, S. Wang, F. Li, and S. Yan, "CTBP-Net: Lung nodule segmentation model based on the cross-transformer and bidirectional pyramid," *Biomed. Signal Process. Control*, vol. 82, p. 104528, 2023.
- [14] C. S. R. Annavarapu, S. A. B. Parisapogu, N. V. Keetha, P. K. Donta, and G. Rajita, "A Bi-FPN-based encoder-decoder model for lung nodule image segmentation," *Diagnostics (Basel)*, vol. 13, no. 8, p. 1406, 2023.
- [15] R. Mothkur and B. N. Veerappa, "Classification of lung cancer using lightweight deep neural networks," *Proc. Comput. Sci.*, vol. 218, pp. 1869–1877, 2023.
- [16] S. Tyagi, D. T. Kushnure, and S. N. Talbar, "An amalgamation of vision transformer with convolutional neural network for automatic lung tumor segmentation," *Comput. Med. Imaging Graph.*, vol. 108, p. 102258, 2023.
- [17] M. Wahengbam and M. Sriram, "An MRI lung tumor segmentation and classification using neural networks," In *International Conference on Communication, Electronics and Digital Technology*. Singapore: Springer Nature Singapore, February 2023, pp. 605–616.
- [18] M. Nagabushanam et al., "Detection and localization of brain tumors using fractional Hartley transform and adaptive neuro-fuzzy inference system classification methods," *J. Ambient Intell. Hum. Comput.*, vol. 14, pp. 8851–8858, 2023.
- [19] A. Alameen, "Smart lung tumor prediction using dual graph convolutional neural network," *Intell. Autom. Soft Comput.*, vol. 36, no. 1, pp. 369–383, 2023.
- [20] M. A. Javed, H. Bin Liaqat, T. Meraj, A. Alotaibi, and M. Alshammari, "Identification and classification of lungs focal opacity using CNN segmentation and optimal feature selection," *Comput. Intell. Neurosci.*, vol. 2023, p. 6357252, 2023.
- [21] G. Aresta, C. Jacobs, T. Araújo, A. Cunha, I. Ramos, B. van Ginneken, and A. Campilho, "iW-Net: an automatic and minimalistic interactive lung nodule segmentation deep network," *Sci. Rep.*, vol. 9, no. 1, p. 11591, 2019.
- [22] Y. H. Ali, V. S. Chooralil, K. Balasubramanian, R. R. Manyam, S. K. Raju, A. T. Sadiq, and A. K. Farhan, "Optimization system based on convolutional neural network and internet of medical things for early diagnosis of lung cancer," *Bioeng. (Basel)*, vol. 10, no. 3, p. 320, 2023.
- [23] G. S. Nandeesh, M. Nagabushanam, S. P. Kumar, and S. Nandini, "Lung parenchyma segmentation and nodule detection using deep learning," *J. Opt.*, vol. 53, no. 1, pp. 635–642, 2023.
- [24] H. Seo, C. Huang, M. Bassenne, R. Xiao, and L. Xing, "Modified U-Net (mU-Net) with incorporation of object-dependent high level features for improved liver and liver-tumor segmentation in CT images," *IEEE Trans. Med. Imaging*, vol. 39, no. 5, pp. 1316–1325, 2019.
- [25] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 2011–2023, 2020.
- [26] H. Seo, C. Huang, M. Bassenne, R. Xiao, and L. Xing, "Modified U-Net (mU-Net) with incorporation of object-dependent high level features for improved liver and liver-tumor segmentation in CT images," *IEEE Trans. Med. Imaging*, vol. 39, no. 5, pp. 1316–1325, 2019.
- [27] D. Amali and M. Dinakaran, "Wildebeest herd optimization: a new global optimization algorithm inspired by wildebeest herding behaviour," *J. Intell. Fuzzy Syst.*, vol. 37, no. 6, pp. 8063–8076, 2019.