



# Automated biomedical image classification using multi-scale dense dilated semi-supervised u-net with cnn architecture

Anoop V<sup>1</sup> · Bipin PR<sup>2</sup> · Anoop BK<sup>3</sup>

Received: 8 October 2022 / Revised: 24 July 2023 / Accepted: 23 August 2023 /  
Published online: 8 September 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

## Abstract

Biomedical research heavily relies on automated image classification to enhance understanding of protein structure and function. This study proposes a novel approach for automating biomedical image categorization, addressing the challenges posed by intricate geometric correlations among various categorical biological patterns. The proposed model incorporates modified histogram normalization (MHN) for image pre-processing, utilizing bi-cubic interpolation and a high boost filter to enhance image resolution and contrast. For segmentation tasks, the study introduces the multi-scale dense dilated semi-supervised U-Net (MDSSU-Net), which combines a convolution block and skip connection path within an enhanced encoder-decoder framework. The semi-supervised U-Net approach allows us to train the model with limited labelled data, significantly reducing the need for extensive annotations. To classify cancer cells in textured bio-images, we employ Grey Level Co-occurrence Matrix (GLCM) and Haralick's feature extraction, which describes pixel intensity relationships within images. The task of automatic cancer classification is particularly challenging, considering the numerous histopathological images that require analysis to detect subtle abnormalities. For an accurate evaluation, we utilize performance metrics such as Dice Coefficient (DSC), Mean Intersection Over Union (MIoU), recall, precision, accuracy, sensitivity, specificity, and F1-score. The Elephant Herding Optimization (EHO) method is employed to design a unique convolutional neural network, known as C-Net, for the classification of biological images. The experimental results demonstrate the superiority of the proposed model. The MDSSU-Net segmentation framework showcases improved performance, efficiently handling diverse segmentation challenges. Moreover, the semi-supervised U-Net training approach significantly reduces the dependence on labelled data, enhancing the model's adaptability to various biomedical datasets. In cancer cell classification, the combination of GLCM and Haralick's feature extraction proves highly effective, enabling the automatic detection of cancerous cells with high accuracy. The C-Net design, utilizing EHO optimization, outperforms other network architectures, achieving superior classification results for complex biological images. This automating biomedical image classification, contributes to the broader adoption of automated image analysis in modern biological studies and diagnostic procedures.

**Keywords** Biomedical image · Segmentation · Classification · Haralick · Dense dilated semi-supervised U-Net · Convolutional neural network

## 1 Introduction

Today's healthcare system relies heavily on medical imaging to carry out non-invasive diagnostic treatments. This entails creating functional and illustrative models of the internal organs and systems of the human body for use in clinical analysis. Its varieties include molecular imaging, magnetic resonance imaging (MRI), and ultrasonic imaging. X-ray-based approaches, such as conventional radiography, computed tomography (CT), and mammography, are also included. In addition to these medical imaging modalities, clinical images are increasingly being utilised to diagnose a variety of disorders, particularly those that are skin-related [1, 2]. By separating various tissues, organs, lesions and biological structures, biomedical image segmentation is a current topic of research in medical analysis that can support surgical planning, medical diagnosis, and treatment. Typically, pathologists perform segmentation manually. This procedure is laborious and time-consuming. The price of manual segmentation is rising, along with the quantity and variety of medical images. Therefore, it is essential to automatically segment biomedical images. The complexity of the biomedical objects' structures, as well as the low contrast, noise, and other impacts brought on by different medical imaging modalities and procedures, make this technique difficult [3].

The expense of manual segmentation is increasing in tandem with the growth in the quantity and variety of medical images. Consequently, automatic segmentation of biomedical images has become essential. This task is difficult, nevertheless, for several reasons, including the complicated differences in the structure of the biomedical items and the low contrast, noise, and other effects brought on by different medical imaging modalities and procedures [4, 5]. Low image quality and the absence of standardised annotation techniques for different imaging modalities can affect annotation quality. The quality of annotations can also be affected by other elements, such as the annotator's attention span, display type, image-annotation software, and data misinterpretation due to lightning circumstances. A faster, more accurate and more dependable alternative to manual image segmentation is an automated computer-aided segmentation-based diagnostic system that can transform clinical operations and enhance patient care. Computer-aided diagnosis will ease the burden on experts and lower the overall cost of attention [6, 7].

In biomedical image segmentation, U-Net architectures are typically employed to automate the prevention and analysis of target regions or sub-regions. Deep convolutional neural networks, residual neural networks, and adversarial networks are examples of advanced deep learning approaches. For establishing computer-aided diagnostic tools for the early detection and treatment of illnesses like brain tumours, lung cancer, Alzheimer's, breast cancer, etc., using various modalities, U-Net-based approaches have recently demonstrated performance at the cutting edge in various applications [8, 9]. Multilevel dilated residual blocks were used instead of the convolutional blocks of the conventional U-Net to improve the learning capabilities. To eliminate the semantic gap and recover the data lost when concatenating features from the encoder to decoder units; this also advises adding nonlinear multilevel residual blocks are added to skip connections [10]. However, utilising current biomedical image categorisation techniques and approaches, it is not possible to extract further novel image attributes with unbalanced features. Biological-image categorisation and deep feature extraction have been proposed using a diagonal bilinear interpolated deep residual network [11].

Despite numerous attempts at this tricky quantitative analytic problem, it remains challenging to perform accurate automatic segmentation, particularly for soft tissue

organs. Deep learning-based networks have gained impressive performance in image processing during the past ten years thanks to the increased availability of datasets. An end-to-end multilayer network called RCGA-Net was introduced, inspired by cutting-edge deep learning research. It comprises an encoder-decoder backbone that incorporates a global context extraction module to highlight more important information and a coordinate attention mechanism based on space and channel [12, 13]. A deep bidirectional network was developed using the LMSEr self-organizing network, a groundbreaking model that folds an autoencoder along the central hidden layer so that the neurons on the paired layers between the decoder and encoder merge into one another, forming bidirectional skip connections. Although feedback linkages improve segmentation accuracy, they may also introduce noise [14, 15]. However, the task of automatic segmentation is challenging due to the complex differences in biomedical object structures, low contrast, noise, and other effects from various imaging modalities and procedures. The quality of manual annotations can also be affected by factors such as the annotator's attention span, display type, image-annotation software, and misinterpretation due to lighting conditions, impacting the accuracy of segmentation.

To address this problem, U-Net architectures and advanced deep-learning techniques have been employed in biomedical image segmentation. These approaches have shown promising performance in various applications, including early detection and treatment of diseases like brain tumours, lung cancer, Alzheimer's, and breast cancer, using various modalities. Despite the progress, current biomedical image categorization techniques still struggle to extract novel image attributes with unbalanced features. Moreover, accurate automatic segmentation, particularly for soft tissue organs, remains challenging, even with deep learning-based networks. Feedback linkages in deep bidirectional networks, while improving segmentation accuracy, can also introduce noise, creating technical gaps in achieving precise and reliable automated segmentation.

In light of these technical gaps, the current study aims to address the challenges associated with accurate and efficient automated biomedical image segmentation. By proposing novel approaches and leveraging advanced deep learning techniques, the study seeks to overcome the limitations of current methods and contribute to the development of more reliable computer-aided diagnostic tools in medical research and patient care.

### **Motivation of the study**

- Biomedical research heavily relies on accurate and efficient image classification for understanding complex protein structures and functions.
- Manual image analysis is time-consuming and labour-intensive, necessitating the development of automated image classification methods.
- Automating image categorization can significantly reduce the human workload and expedite research processes in the biomedical field.
- The challenges posed by intricate geometric correlations among various biological patterns call for advanced and specialized image classification techniques.

### **Contribution of the Current Study**

- Introducing a novel approach for automated biomedical image categorization, addressing the complexities of intricate geometric correlations.
- Proposing the use of modified histogram normalization (MHN) for effective image pre-processing, enhancing resolution and contrast.

- Introducing the multi-scale dense dilated semi-supervised U-Net (MDSSU-Net) for segmentation tasks, with improved performance in diverse scenarios.
- Implementing a semi-supervised U-Net training approach, reducing the need for extensive labelled data and enhancing model adaptability.
- Leveraging Grey Level Co-occurrence Matrix (GLCM) and Haralick's feature extraction for accurate classification of cancer cells in textured bio-images.
- Designing a unique convolutional neural network, C-Net, using the Elephant Herding Optimization (EHO) method, achieving superior results for biological image classification.

The main objective of this research is to develop an advanced and efficient automated image classification system for biomedical applications. The proposed model aims to enhance the accuracy and reliability of image analysis, reduce manual workload, and accelerate discoveries in protein structure understanding and cancer cell classification. The article is organized as follows to present a clear and coherent exploration of the proposed model and its outcomes. A comprehensive review of existing methodologies and techniques related to biomedical image classification is presented in Section 2. Section 3 includes the problem definition and motivation, and Section 4 details the pre-processing materials and methods of the proposed technique. The outcomes of the proposed model are thoroughly evaluated in Section 5, comprehensive discussion of the experimental findings and the superiority of the proposed model is provided in Section 6. The research concludes by summarizing the achievements and contributions of the proposed model in automating biomedical image classification in Section 7.

## 2 Literature survey

Biomedical signal processing, biomedical engineering, gene analysis, and biomedical image processing are only a few subfields of biomedical science research. Classification, detection, and recognition are of tremendous importance in the investigation and diagnosis of diseases. Das et al. [16] developed a biomedical image classification technique. Brain magnetic resonance imaging was used to analyse the brain tumour, and chest X-rays that were influenced by COVID were classified using an ensemble approach. Convolutional neural networks (CNN), long short-term memory, recurrent neural networks, and gated recurrent units are four heterogeneous base classifiers that are considered for this task, and metadata is generated. The ensemble output of the base classifiers, expressed in terms of class probability and labels, was fed into the fuzzy model. While it achieved good results in brain tumour and COVID-19 chest X-ray classification, the ensemble model's interpretability and computational complexity may pose challenges. Alnabhan et al. [17] used the development of biomedical image segregation to simplify complicated CNN hyperparameter associations. The type and size of the kernel, batch size, learning rate, momentum, convolution layer, activation function, and dropout are some of the intricate characteristics of a CNN. For segmentation, the metaheuristic optimisation method of the EVO technique was applied. Although metaheuristic optimization methods can improve segmentation accuracy, they may require extensive computational resources.

Cancer is the leading cause of mortality worldwide. For this crippling condition to be treated properly, early detection is essential. To classify biological images, Barzekara and Yu et al. [18] suggested a novel CNN architecture called C-Net, which comprises

the concatenation of several networks. On two open datasets, BreakHis and Osteosarcoma, C-Net was used to classify histological images. The performance of the model was assessed using various reliability evaluation metrics. However, further investigation is needed to assess its generalizability and performance on diverse datasets. Textural characteristics, clustering, and orthogonal transformations were utilised to identify and isolate breast tumours in a woman in an image. The results of their software implementation on biomedical images of oncological pathologies of breast cancer are the Hadamard transform, oblique transform, discrete cosine transform, Daubechies transform, and Legendre transforms methods of analysis of texture images of breast cancer carried out by Orazayeva *et al* [19]. While these methods can identify breast cancer, their effectiveness on other types of cancer and scalability to larger datasets need evaluation.

CNNs have been the driving force behind deep learning, which has achieved outstanding results in a variety of medical imaging tasks, including image classification, image segmentation, and image synthesis. However, to validate the model and make it interpretable, we also need to know how confident it is in its predictions. To segment images of brain tumours, Sagar *et al.* [20] developed an encoder-decoder architecture based on variational inference methods. While utilising a systematic Bayesian strategy to consider both epistemic and aleatoric uncertainties, this model may segment brain tumours. To improve the capacity of the network to capture intricate pixel correlations, the number of learnable parameters must be increased, which frequently results in overfitting or weak resilience. However, the model's complexity and training requirements may limit its application in resource-constrained environments. Hyperconvolution, which was proposed by Ma *et al.* [21], is a potent new building block that implicitly depicts the convolution kernel as a function of kernel coordinates. It was centred on difficult biomedical image segmentation tasks, and substituting hyper convolutions for standard convolutions results in more effective designs that yield higher accuracy. Nevertheless, the trade-off between increased model capacity and overfitting needs to be carefully managed. The first attempt to attain a sublinear runtime and constant memory complexity for the random walker algorithm was made by Drees *et al.* [22], who presented a hierarchical design. The technique was quantitatively assessed using fictitious data and the CT-ORG dataset, where the anticipated algorithmic benefits were found. These models must be trained using many trainable parameters, floating-point operations per second, and powerful computing resources. Real-time semantic segmentation in low-powered devices is highly challenging due to these characteristics. However, its performance on real-time semantic segmentation tasks in low-powered devices needs to be further explored.

As a result, Olimov *et al.* [23] changed the U-Net model in their study by creating a fast U-Net (FU-Net) that relies on bottleneck convolution layers in the model's contraction and expansion paths. By providing a cutting-edge performance, the proposed model can be used in semantic segmentation applications. Using biomedical images, an intelligent method for classifying and detecting oral squamous cell cancer is developed by Alanazi *et al* [24]. Additionally, the DBN model based on the extended grasshopper optimization algorithm (EGOA) is used for the identification and classification of oral cancer. However, further validation on larger and more diverse datasets is necessary. The EGOA algorithm is used to tune the DBN model's hyperparameters, improving the classification results. Hamza *et al* [25] developed a perfect deep transfer attempting to learn a human-centric biomedical diagnosis model for the detection of acute lymphoblastic leukaemia. The proposed model's main objective is to recognise and categorise acute lymphoblastic leukaemia from blood smear images. Additionally, it divides the images into pieces using an MFCM algorithm. Numerous simulations were run on an open-access dataset to examine the improved

performance of the method. However, the model's performance on other types of cancer and its interpretability require further investigation. Mandakini Behera et al. [26] proposed the firefly algorithm can be integrated with the popular particle swarm optimization algorithm. In this paper, two modified firefly algorithms, namely the crazy firefly algorithm and variable step size firefly algorithm, are hybridized individually with a standard particle swarm optimization algorithm and applied in the domain of clustering. The results obtained by the two planned hybrid algorithms have been compared with the existing hybridized firefly particle swarm optimization algorithm utilizing ten UCI Machine Learning Repository datasets and eight Shape sets for performance evaluation. In addition to this, two clustering validity measures, Compact-separated and David–Bouldin, have been used for analyzing the efficiency of these algorithms. The experimental results show that the two proposed hybrid algorithms outperform the existing hybrid firefly particle swarm optimization algorithm. However, the comparison with other optimization algorithms and scalability to larger datasets should be considered.

Sirisati Ranga Swamy et al. [27] predicted the effects of the disease outbreak and help detect the effects in the coming days. In this paper, Multi-Features Decease Analysis (MFDA) is used with different ensemble classifiers to diagnose the disease's impact with the help of Computed Tomography (CT) scan images. There are various features associated with chest CT images, which help know the possibility of an individual being affected and how COVID-19 will affect the persons suffering from pneumonia. The current study attempts to increase the precision of the diagnosis model by evaluating various feature sets and choosing the best combination for better results. The model's performance is assessed using Receiver Operating Characteristic (ROC) curve, the Root Mean Square Error (RMSE), and the Confusion Matrix. It is observed from the resultant outcome that the performance of the proposed model has exhibited better efficient. However, further evaluation of diverse CT scan datasets and comparison with other forecasting models is essential.

In this work, the article aims to address some of the limitations observed in the existing methods by proposing a novel approach that incorporates modified histogram normalization, the MDSSU-Net segmentation framework, and the Elephant Herding Optimization (EHO) method for convolutional neural network design. This model is expected to offer improved performance, interpretability, and efficiency in biomedical image classification and segmentation tasks, contributing to the broader adoption of automated image analysis in biomedical research and diagnostic procedures. The experimental results will be discussed in detail, including comparisons with other well-known evaluation metrics, to validate the effectiveness of the proposed approach.

### 3 Research problem definition and motivation

Today's healthcare system relies heavily on medical imaging to carry out noninvasive diagnostic treatments. This entails creating functional and illustrative models of the internal organs and systems of the human body for use in clinical analysis. It comes in a variety of forms, including X-ray-based techniques including traditional X-rays, computed tomography (CT), and mammography; molecular imaging; magnetic resonance imaging (MRI); and ultrasonic imaging. Clinical images, in addition to these medical imaging methods, are increasingly being utilized to diagnose a variety of disorders, particularly skin-related ones. Medical imaging consists of two different processes: image production and reconstruction, and image processing and analysis. While image analysis collects numerical data or a set

of features from the image for object identification or classification, image processing uses techniques to improve image attributes like noise reduction. Technology advancements have made it easier to acquire photographs, which has resulted in the production of enormous numbers of high-resolution images at extremely low costs. The creation of biological image processing algorithms has significantly advanced as a result. In turn, this has made it possible to create automated algorithms for information extraction through image analysis or evaluation. Diagnostics, healthcare, and drug treatment are just a few of the many uses for biomedical images. Biomedical imaging techniques like microscopy can expose regions and things that are beyond the range of the human eye's normal resolution, exposing significant insights into the structures of the smallest objects. Despite the enormous advantages that biomedical images offer for clinical practice and biomedical research, traditionally only trained researchers and clinicians may assess these images. Visual inspection is susceptible to difficulties from both subjective and objective viewpoints, even when specialists are well-trained to spot characteristic patterns in the photographs, such as odd shapes and colours. When modifying and training a neural network, many expert decisions must be made, including the precise network architecture, the training schedule, and the methods for data augmentation or post-processing. Biomedical image segmentation has long been one of the most crucial tasks in biomedical imaging research due to the widespread use of medical imaging in healthcare.

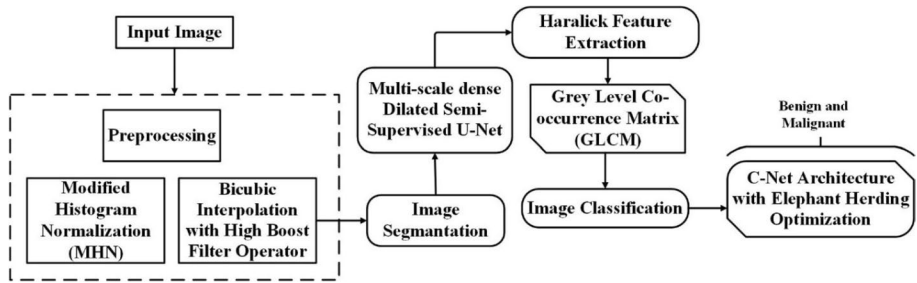
- The need for automated image classification in modern biomedical research to accelerate understanding and discovery.
- Enhancing the efficiency of image analysis to handle the vast amount of data generated in the biomedical field.
- Overcoming the challenges of intricate geometric correlations and subtle abnormalities in biomedical images.
- Advancing the state-of-the-art in image segmentation and classification, leading to more reliable and accurate results.
- Offering a robust and adaptable model that can be applied to various biomedical datasets with limited labelled data.

By addressing these motivations and making significant contributions, the current study aims to revolutionize automated biomedical image classification and facilitate breakthroughs in the field of biological research and diagnostics.

## 4 Proposed research methodology

In many countries, cancer is the leading cause of mortality. For this crippling condition to be treated properly, early detection is essential. It is difficult to automatically classify the kind of cancer since pathologists must analyse many histopathological images to find minute abnormalities. As a result, the study suggested a segmentation and classification technique for biomedical images utilising a collection of brain images. The suggested work's flow diagram is shown in Fig. 1.

The data is subjected to a specific set of operations during the data pre-processing stage, including resizing and normalisation to lessen intensity variation in image samples, augmentation to produce more training samples to prevent the problem of class biases and overfitting, removal of pointless noise or artefacts from the data samples, etc. employing



**Fig. 1** Flow of the Proposed Work

modified histogram normalisation (MHN) and bicubic interpolation with high boost filter operation to pre-process the image. Computerized medical image analysis requires biomedical image segmentation. Positive outcomes have been obtained with the U-Net and its variations across many datasets. Considering this, the research suggested the multi-scale dense dilated semi-supervised U-Net (MDSSU-Net), in which at each resolution level, the model is supervised and the field of view in each level varies based on the depth of the resolution layer. Therefore, Haralick features are used in the feature extraction procedure. Haralick characteristics were extracted using a grey-level co-occurrence matrix (GLCM). A timely cancer diagnosis is essential to receiving effective therapy for this crippling condition. It is difficult to automatically classify the kind of cancer since pathologists must analyse many histopathological images to find minute abnormalities. In this study, a CNN architecture with an elephant herding optimization technique, known as C-Net, is proposed to categorise biomedical images.

#### 4.1 Image dataset

Two publicly accessible databases provided the MRI datasets used in the investigation. Strong magnetic fields are utilised in MRI imaging to create images of bodily tissues, organs, and physiological processes. Soft tissues or non-bony body parts are scanned using MRI technology. MRI scans can be used to discriminate between grey and white matter in the brain, which helps doctors identify tumours and aneurysms. The Open Access Series of Imaging Studies (OASIS) project has compiled neuroimaging datasets with more than 2000 MRI sessions for biomedical imaging researchers.

#### 4.2 Preprocessing image

The scarcity of massive data for training new machine/deep learning models poses significant issues in the field of medical imaging because these models are inherently data-hungry. Data augmentation offers a means to overcome the problems with tiny datasets by creating fake data and adding it to the training set. To achieve this, all experiment-related image intensities are initially scaled to the range of 0 to 1. The pre-processing algorithm receives the raw images. The two main modules of the proposed pre-processing approach are artefact and noise removal and resampling and normalisation. These photos were also utilised to train the suggested models and to verify the suggested normalising and smoothing techniques. The Modified Histogram Normalization (MHN) technique was used to



handle image normalisation. Bicubic interpolation is utilised in pre-processing to denoise captured images, and a high boost filter is employed to increase their contrast.

#### 4.2.1 Modified Histogram Normalization

In the suggested MHN approach, intensity scaling and normalisation are the two processes  $q_{\min}$  and  $q_{\max}$  that are involved. First, during intensity scaling, the region of interest with low intensity ( $ROI_l$ ) and the region of interest with high intensity are represented by ( $ROI_h$ ) in the original (reference high-quality) image and are determined from its histogram after smoothing, assuming that the minimum and maximum intensity levels on the standard scale are denoted by and. A function  $g(x, y)$  shown in Eq. (1) is used to transfer the image's intensity to values between  $ROI_h$  and  $ROI_l$ .

$$g(x, y) = \frac{f(x, y) - ROI_l}{ROI_h - ROI_l} \quad (1)$$

Where  $f(x, y)$  denotes the equivalent transformed grayscale value  $x, y$  and  $g(x, y)$  is the corresponding grayscale value of the original image at coordinate. Second, using a function  $h(x, y)$  depicted in Eq. (2), the original image is shifted and extended during normalisation to cover all the grayscale levels in the image.

$$h(x, y) = \frac{ROI_h - ROI_l}{q_{\max} - q_{\min}} (g(x, y) - q_{\min}) + ROI_l \quad (2)$$

The original image can be scaled up between the lower boundary  $m_1$  and upper boundary  $m_2$  if the target histogram of the original image functions  $g(x, y)$  at  $q_{\min}$  and  $q_{\max}$  spreads up to grayscale levels in the intensity region of interest (ROI). This will cause the pixels in the normalized image to lie between the minimum intensity level ( $ROI_l$ ) and the maximum intensity level ( $ROI_h$ ). The lower boundary and the higher boundary of the original image before scaling are represented by the variables  $m_1$  and  $m_2$ .

Two different linear mappings can be made to normalize the image.  $P_1, \mu_i$  to  $S_1, \mu_s$  and  $\mu_i, P_2$  to  $\mu_s, S_2$  are the first and second mappings, respectively. The standard scale's lower and upper bounds are then applied to  $S'_1$ , and  $S'_2$ , respectively, by mapping  $(m_1, P_1)$  to  $(S'_1, S_1)$  and  $(S'_2, m_2)$  to  $(S_2, S'_2)$ . The conversion of the input image's intensity ( $S'_1, S'_2$ ) to  $(m_1, m_2)$  of the standard scale is known as normalization. Equation (3) defines the normalizing function as  $N(x, y)$ .

$$N(x, y) = \left[ \begin{array}{l} \mu_s + (g(x, y) - \mu_i) \frac{S - \mu_s}{P - \mu_i} \\ \mu + (g(x, y) - \mu_i) \frac{S - \mu_s}{P - \mu_i} \end{array} \right] \quad (3)$$

Where  $\mu_i$  and  $\mu_s$  are the mean values for the input image histogram and original histogram, respectively, and  $[\bullet]$  signify the "ceiling" operator.  $P_1$  and  $P_2$  come from the supplied image's pixel values. The low-quality photos (input images) were then normalised across the  $ROI_l - ROI_h$  ranges of their intensity using the high-quality (reference) image for each group. To ensure that all images have the same intensity normalisation, the process is repeated across groups. The MHN will apply local contrast enhancement and edge enhancement to each region of the supplied input image. The HB bi-Cubic method will be loaded with the output of the MHN.

#### 4.2.2 Bicubic interpolation with high boost filter operator

When the processing speed is unimportant, bicubic interpolation is typically used in place of bilinear interpolation or nearest neighbour. A cubic interpolation algorithm called BC uses luminance data from a bicubic array of  $4 \times 4$  pixels to resample the luminance value of a point (16 pixels). To limit the impact of the high-frequency component on the image, the high-frequency component is multiplied by a coefficient. The boost coefficient is the name of this parameter. The original signal component and the high-frequency augmentation component make up the two sections of the high-boost filter. To reduce the high-frequency component's impact on the image, the high-frequency element is multiplied by a coefficient. The boost coefficient is the term given to this coefficient. The following equation can be used to express the high boost filtering of any signal  $f$ .

$$f_b = f + \alpha \cdot HP \otimes f \quad (4)$$

Where,  $\otimes$  indicates the convolution operator,  $\alpha$  denotes the boost coefficient,  $HP$  is the high pass filter,  $f_b$  defines the high boosted signal, and  $\alpha$  is the boost coefficient. The Laplacian operator ( $\nabla^2$ ) specifies a high pass filter that can be used in image processing. Consequently, the high boost filter can be expressed as

$$f_b = f + \alpha \nabla^2 f \quad (5)$$

For interpolation-based resampling, the third-order or cubic polynomial function has typically been used. A cubic polynomial function's generic form is represented as

$$f(x) = (a + bx + cx^2 + dx^3) \quad (6)$$

Where,  $a$ ,  $b$ ,  $c$  and  $d$  denotes the two random constants. As a result, the cubic polynomial function's high boost filtering can be represented as

$$f(x) = (a + bx + cx^2 + dx^3) + \alpha(2c + 6dx) \quad (7)$$

Where,  $\alpha$  is typically taken to be between  $-0.5$  and  $-0.75$ , and  $x$  represents the grey level value. Based on high boost filtering, which altered the high pass term using the Laplacian operator, this approach was used. The third-order polynomial was used to approach the interpolation formula. Consequently, the preceding is a formulation for the High Boost Cubic (HBC) interpolation.

$$y_b(x, \alpha) = B_{-1}y_{i-1} + B_0y_i + B_1y_{i+1} + B_2y_{i+2} \quad (8)$$

By dividing the output size by the original image size, one can calculate the magnification. The boost coefficient, which increases the output image's sharpness, is the second parameter. It is used to multiply with the high-frequency component to reduce its influence. The HBC interpolation technique will compute both parameters to remove noise, and blur, and create high sharpness image results, improving the image resolution.

#### 4.3 Multi-scale dense dilated semi-supervised u-net for image segmentation

Based on U-Net, some medical image segmentation techniques have been created extremely quickly for performance enhancement. Application range, feature

augmentation, training speed optimization, training accuracy, feature fusion, small sample training set, and generalisation improvement are all areas where U-Net is enhanced. Different network structures have been designed using various methods to handle various segmentation issues. It is suggested to use an enhanced encoder-decoder-based segmentation architecture that incorporates a multi-scale densely connected convolution block and skip connection path. Additionally, a semi-supervised U-Net approach is used during the model training process to train the model, which can reduce the need for labelled training data.

The objective of a dense block is to learn all of the  $f$  features from the input features. Additional features  $k$  are learned at each step as the process iterates through numerous steps. Prior convolutional layers are connected to all succeeding layers by channel-wise concatenation, which is the main characteristic of dense connectivity. Based on the initial input and other attributes learnt in earlier layers, each subsequent step learns new features. As a result, there is no longer a need to master redundant features, and studying a wider variety of features is encouraged. The dense block was changed in this work to use both regular and dilated convolutions. Each phase involves learning some  $k/2$  features using a conventional convolution and learning the rest of the features using dilated convolutions with a dilation rate  $r$ . To reduce the possibility of gridding artefacts caused by employing only dilated convolutions, this combination was chosen. Dilated convolutions can effectively learn global context due to their broader receptive field. Standard convolutions, on the other hand, may effectively learn local context because of their denser receptive area.

#### 4.3.1 Multiscale dense dilated U-Net

An “encoder-decoder-refinement” framework governs the DD-Net. The encoder is a down-sampling route that is used to increase receptive fields while reducing the resolution of feature maps. The distinctive usage of dense dilation blocks to capitalise on the advantages of dilated convolutions and dense connectivity is the primary innovation of the DD-Net. Over a two-dimensional feature map  $x$ , dilated convolution is exploited, and the kernel size  $k$  and filter ware applied to each position  $i$  on the output  $y$  as described in

$$y[i] = \sum_k x[i + d \cdot k]w[k] \quad (9)$$

Where the dilation rate  $d$  is equal to the sampling stride of the input signal. Convoluting the input  $x$  with up-sampled filters, which are created by adding  $(d-1)$  zeros between two consecutive filter values along each spatial dimension, is similar to this method. Two convolutional layers and a max-pooling layer make up a standard U-Net block, and the output feature map is transmitted through both layers. The dense dilated U-Net architecture is shown in Figs. 2 and 4.

In the proposed multi-scale architecture, the first convolutional layer of each block has three sets of kernels, each of which operates at a different scale. In other words, one set of kernels operates at the block’s input resolution, another set has a dilation and stride of two to reduce the resolution to half, and a third set has a dilation and stride of four to reduce the resolution to a fourth of the input. The article employed four dense blocks in this architecture’s encoder path, each of which is made up of two CONV (3×3). The batch normalisation (BN) procedure and activation function have been used to add all of the convolutional layers to the proposed network (ReLU). The first node’s output characteristics in the encoder section are described as

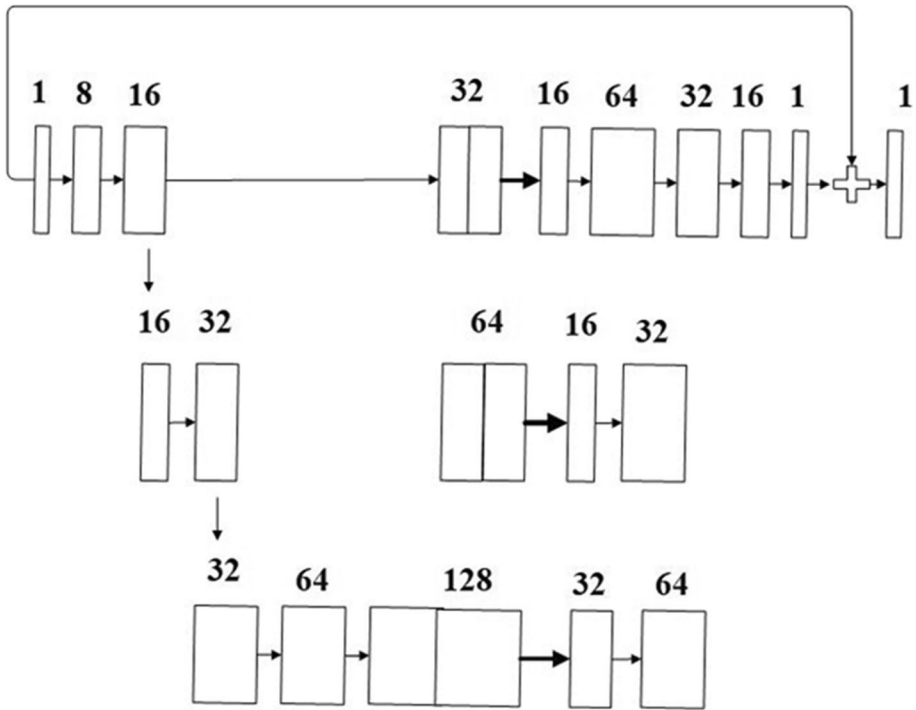


Fig. 2 Dense Dilated U-Net Architectures

$$\chi_{1(U)}^i = C^{4-i} [\chi_1^k]_{k=1}^{i-1}, i = 1, 3 \tag{10}$$

The first encoder node of the U-first net’s *i*th output feature is denoted by  $\chi_{1(U)}^i$  which is stated in the equation above.  $[\cdot]$  identifies the concatenate function and  $C^r$  represents the dilated convolution with a dilation rate equal to *r*. The model only used the max-pooling function to generate the input features for the encoder node; after pooling, the size of the input features is reduced by a factor of two. Since max pooling layers frequently produce high-frequency material that could lead to gridding artefacts, they were eliminated from the DD-Net. In their place, two-strategy  $2 \times 2$  convolutional layers were added, enabling the CNN to learn a more beneficial transformation for spatial down-sampling. The end of the initial network was also extended with a shallow “refinement” network made up of a dense dilation block and two  $3 \times 3$  convolutional layers. The CNN can further hone the image and eliminate artefacts at the maximum spatial resolution thanks to these extra layers. The final segmentation map is created using a Softmax layer as a logistic function.

### 4.3.2 Semi-supervised segmentation

The method uses both unlabelled and labelled data during training, following the mainstream of semi-supervised segmentation approaches. The study’s labelled dataset  $D_l$  of image-label pairs  $(x, y)$ , which includes both image  $x \in R^\Omega$  and ground-truth labels  $y \in \{1, \dots, C\}^\Omega$ , and a larger unlabelled dataset  $D_u$ , which consists of images without their annotations. These are used to explore a semi-supervised segmentation job. In this case,

C is the number of segmentation classes, and  $\Omega = \{1, \dots, W\} \times \{1, \dots, H\}$  denotes the image space (i.e., set of pixels). The network's parameters are discovered by maximising the following loss function.

$$L(\theta; D_i; D_u) = L_{SP}(\theta, D_l) + \lambda_1 L_c(\theta, D_u) \quad (11)$$

This loss is made up of four distinct terms, each of which relates to a different segmentation feature and whose relative weight is managed by hyper-parameters  $\lambda_1, \lambda_2, \lambda_3 \geq 0$ . The  $L_{SP}$  uses labelled data, just like in conventional supervised approaches, and  $D_l$  imposes the network's pixel-by-pixel prediction for an annotated image to resemble the ground truth labels. The technique leverages the well-known cross-entropy loss, while other segmentation losses, such as the Dice loss, could also be taken into consideration.

$$L_{SP}(\theta, D_l) = -\frac{1}{|D_l| |\Omega|} \sum_{(x,y) \in D_l} \sum_{(i,j) \in \Omega} y_{ij} \log f_{ij}(x, \theta) \quad (12)$$

The model utilizes this unlabelled data to regularise learning and direct the optimization process toward optimal solutions because it lacks annotations for the images  $D_u$  in the dataset.

#### 4.4 Feature extraction based on haralick features

One of the crucial tasks that improve the effectiveness of the entire system is feature extraction. The feature explains the computational property of the input image. In this study, textured bio-images of cancer cells are classified using Haralick's features-based GLCM. Haralick's symmetric Grey Level Co-occurrence Matrix-derived textural features (GLCM). These features describe the relationship between the intensities of two pixels in an image that are separated by a specific amount and facing a specific direction. The link between the intensities of nearby pixels can be ascertained using these properties. Information on the spatial distribution of tone and texture variations in an image is contained in the relationship between pixels. In addition to inter-pixel relationships, GLCM also provides periodicity and spatial dependencies of grey levels. In the suggested study, each GLCM's 24 Haralick characteristics are extracted for each image. The angular second moment, contrast, entropy, correlation, variance, homogeneity, sum average, sum entropy, sum variance, square difference, HX and HY entropies, Inverse Difference Moment, difference variance, difference entropy, maximal correlation coefficient, and information correlation are among the features.

##### 4.4.1 Gray-level co-occurrence matrix (GLCM)

This method has been extensively applied in image analysis applications, particularly in the biomedical industry. To extract features, there are two processes. The GLCM is computed in the first phase, and the second step computes the texture characteristics based on the GLCM. According to the grey level, GLCM displays the frequency of each grey level at a pixel concerning each other at a fixed geometric point. In this study, the horizontal direction 00 with the nearest neighbour range of 1 was employed. A grey-level transition was made to two pixels in an input image to apply the GLCM characteristics. The GLCM characteristics must be computed in two steps. The first step is to segregate the pair-wise spatial co-occurrence of image pixels by distance  $d$  and

direction angle  $\theta$ . For nearby and reference pixels, a spatial link is established between the two pixels. In the second stage, scalar quantities that make use of the representation of various characteristics of an image are used to compute GLCM features. This procedure creates the grey-level co-occurrence matrix, which includes some grey-level pixel pairings in the target image or selected area of the target image.

**Input:** Load  $I$  with input images  
**Output:** 24 features  
 Consider each image in  $I$  with pixel  $P_d$  in  $(i, j)$   
**Do for**  
     Compute the GLCM of the current pixel  $P_d$  in a neighbourhood of size  $M \times M$   
     Calculate Pixel probability  $P(i, j, d, \theta)$  having grey-levels  $i$  and  $j$ .  
     Extract Haralick features of each GLCM matrix in  $M$   
**end for**  
**return**  $M$

**Algorithm 1** Pseudocode of Haralick Feature Extraction

The pseudocode for the Haralick feature extraction is shown in Table 1. The resulting GLCM matrix is  $M \times M$ , where  $M$  stands for the image's grayscale values and  $P_d$  for the matrix's spatial distributions of nearby pixels. Consider the pixel probability  $P(i, j, d, \theta)$ , which represents the likelihood that two pixels with the corresponding grayscales  $i$  and  $j$  are located a certain distance apart.  $P_d(i, j)$  consequently, indicates the  $(i, j)$ th elements of the GLCM matrix.

#### 4.5 C-Net architecture for image classification

The Outer, Middle, and Inner networks make up C-three Net's primary sections. Four CNN's taken from VGG19 make up the outer one. Each of the outer networks has the following architecture: The images are concurrently fed into the input layer of each outer network, followed by many convolutional layers, a max-pooling layer, and finally the first block. A combination of numerous convolutional layers followed by a pooling layer is referred to as a block in this context. The first block has 64 filters with a size of  $3 \times 3$  with the same padding, while the max-pooling filters have a size of 22 with a stride of two. For three additional blocks, the same structure is repeated in the same manner, with the exception that the final block is not multiplied by 2, but the number of filters is. The max-pooling layer has been removed for the final block to avoid additional

**Table 1** Simulation System Configuration

MATLAB	Version R2021a
Operation System	Windows 10 Home
Memory Capacity	6GB DDR3
Processor	Intel Core i5 @ 3.5GHz
Simulation Time	10.190 seconds

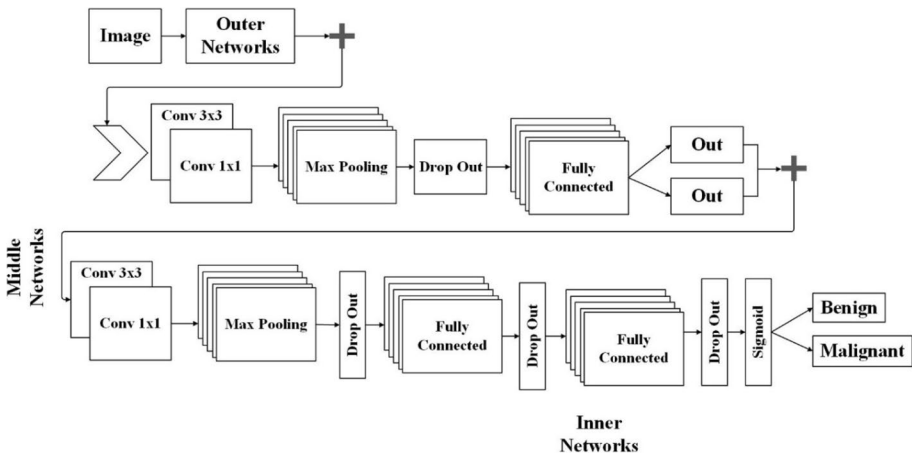


Fig. 3 C-Net Architecture with Outer, Middle and Inner Networks

output reduction. As the activation function, Rectified Linear Unite (ReLU) is applied to the convolutional layers. The C-Net architecture can be seen in Figs. 3 and 5.

One of the crucial procedures in the C-Net model stated in Eq. 13 is the Concatenation of the returning output features, as indicated by the sign  $\oplus$  in Figs. 3 and 5. Each output from the outside networks is subjected to this process twice.

$$f(y, m) = \left( (y_{m,n,c_i}) \oplus (w_{m,n,c_j}) \right) = Z_{m,n,c_i+c_j} \tag{13}$$

Where,  $(m, n)$  signifies the dimensions of the outputs of the networks,  $y$  and  $w$  states the feature maps of the various networks, the number of channels on each output is denoted by  $(c_i, c_j)$ , the concatenation  $\oplus$  with regard to the axis and channel of the feature map. Finally, the output of the concatenation operation  $Y$ , would be the input for the middle networks. The Middle networks’ input comes from features retrieved from the Outer networks. It comprises four convolutional layers stacked on top of one another, each with 256 filters and a filter size of  $3 \times 3$ . The previous convolutions have been followed with an 11 convolution to compress the feature maps and get around the model’s complexity. After the 11 convolutional layers, a max-pooling layer with a stride of 2 has been added to finish the first block of the Middle networks. In the second block of the Middle networks, the same construction is replicated. As shown in Eq. 1, the outputs (feature maps) from the Middle networks will be combined and used as the input for the Inner network. This ensures that each network has the max-pooling layer to produce effective feature descriptors. Drop out, a regularisation strategy that involves randomly shutting off some layer units has the extreme consequence of preventing the network from overfitting. Every block of the Middle networks has received it.

The features supplied by the Middle networks are then used as input by the Inner network. Only one block, with two convolutional layers, a filter size of  $3 \times 3$ , a stride of 1, the same padding, and 256 filters, is present in the inner network. An 11-convolutional layer with identical configurations and a max-pooling layer with a size of  $2 \times 2$  and a stride of 2 are also included in the study. The maximum pooling layer that the Inner

network has returned is flattened out into a vector that is connected to two FC layers, each of which has 1024 units. On both of the FC levels, dropout has been applied. At the output layer with two nodes, benign and malignant, as shown in Figs. 3 and 5, a sigmoid is employed as the activation function, as shown in Eq. 2.

$$z = w^T X + B \quad (14)$$

$$\hat{y} = \text{Sig}(z) = \frac{e^z}{e^z + 1} \quad (15)$$

Where  $B$  is the bias and  $z$  represents the dot product of filter  $w$  with a portion of the image that is the same size as the filter. The proposed model's loss function is a cross-entropy function, shown as follows:

$$L(\hat{y}, y) = -\left(\sum_{i=1}^N y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i)\right) \quad (16)$$

Where  $y_i$  is the  $i^{\text{th}}$  element of the model's output  $y$ , and  $\hat{y}_i$  is the  $i^{\text{th}}$  label  $y$  of the  $N$  classes.

The calibration of the convolutional network hyperparameters must produce highly accurate results to achieve high levels of accuracy in image classification, and this operation consumes a significant amount of computational time and resources.

#### 4.5.1 Elephant herding optimization

To achieve high accuracy in image classification, convolutional network hyperparameter tuning is a crucial issue in this field; yet, this operation requires a lot of computational effort. Current knowledge cannot begin to comprehend the complexity of interactions between all network hyperparameters, including activation type, layer size, number of layers, and connection patterns. The orthogonal experiment, which chooses the ideal arrangement of model parameters, optimises the network hyperparameters of the CNN model. A metaheuristic strategy based on a hybridised version of the elephant herding optimization swarm intelligence metaheuristics has been developed in this scientific research report to automatically search and target the near-optimal values of convolutional neural network hyperparameters. A system for immediate image categorization of benign and malignant from a biological image has been developed using the EHO for convolutional neural network hyperparameter tuning.

Different optimization problems are solved using the EHO swarm-based search method. The elephant herding behaviour served as the inspiration for this programme. Elephants are divided into clans, each of which is headed by a matriarch. The male elephants that are adults depart from their family unit. Therefore, these two elephant group behaviours result in two operators, namely the clan updating operator and the clan separating operator. In the EHO algorithm, the updating operator updates each solution  $j$  in each clan  $ci$  following its current position and matriarch  $ci$ . The population diversity is then improved at later generations of algorithm execution through the separation operator. First, the EHO model's hyperparameters are chosen as the optimization objective, and each particle's position information is initialised at random in the predetermined hyperparameter value space. Additionally, the particles are separated into populations that can adapt.



A vector of integer numbers with a dimension of  $2N$  is used to represent each member of the population, where  $N$  is the number of unidentified sensor nodes. The population is initially split up into  $n$  clans. The updating operator is modelled by shifting each solution's rank within the clan by  $ci$ , the matriarch with the highest fitness value in the generation.

$$x_{n,ci,j} = x_{i,j} + \alpha \times (x_{b,ci} - x_{ci,j}) \times r \quad (17)$$

Where,  $x_{n,ci,j}$  denotes the solution  $j$  new location within clan  $ci$ ,  $x_{ci,j}$  denotes its previous position within clan  $ci$ , and  $x_{b,ci}$  is the best solution within clan  $ci$  that has been discovered. Matriarch  $ci$  influence on  $x_{ci,j}$  is shown by the scale factor  $\alpha \in [0, 1]$ , whereas  $r \in [0, 1]$  is a random variable with a uniform distribution. The fittest solution in each clan  $ci$  is updated using the following expression.

$$x_{n,ci,j} = \beta \times x_{ce,ci} \quad (18)$$

The influence of the  $x_{ce,ci}$  on the updated individual is denoted by  $\beta \in [0, 1]$ . For the  $d$ th dimension problem, the centre of the clan  $ci$  can be determined as

$$x_{ce,ci,d} = \frac{1}{n_{ci}} \times \sum_{j=1}^d x_{ci,j,d} \quad (19)$$

Where  $n_{ci}$  is the number of solutions in clan  $ci$  and  $D$  denotes the dimension of the search space in the  $d$ th dimension in the condition of  $1 < d < D$ . The following is the pseudocode of the corresponding algorithm (algorithm 2). The maximum generation is MaxGen.

**Initialization:** Set population  $N$ , maximum generation  $Maxgen$ , generation counter  $t=1$   
 Divide the population into  $n$  clans  
**While**( $t < Maxgen$ )**do**  
   Sort all solutions according to their fitness  
   **For all** clans  $ci$  **do**  
     **for all** solution  $j$  in the clan  $ci$  **do**  
       Update  $x_{ci,j}$  and generate  $x_{n,ci,j}$  using equation 17  
       Select and retain better solutions between  $x_{ci,j}$  and  $x_{n,ci,j}$   
       Update  $x_{b,ci}$  and generate  $x_{n,ci,j}$  using equation 18  
       Select and retain better solutions between  $x_{b,ci}$  and  $x_{n,ci,j}$   
     **end for**  
   **end for**  
**end for**

**Algorithm 2** Pseudocode for Elephant Herding Optimization.

Every time an algorithm generation is executed, the separating operator is used on the population's worst offender.

$$x_{w,ci} = x_{min} + (x_{max} - x_{min} + 1) \times r \quad (20)$$

Where  $x_{max}$  and  $x_{min}$  stand for the individual's upper and lower bounds of position,  $x_{w,ci}$  denotes the member of clan  $ci$  with the weakest fitness, and  $r \in [0, 1]$  is a uniformly

distributed random number. The mainframe of EHO is outlined under the descriptions of the clan-updating operator and separating operator.

## 5 Experimentation and results discussion

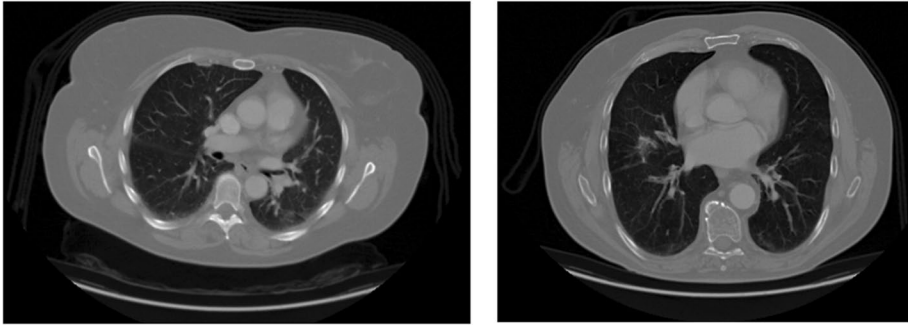
The dataset collection, experimental setup, results, conclusions, and debates are all presented in this part. Also covered here is how well various features perform when used with various machine learning models. The proposed algorithm has been implemented using MATLAB software. Table 1 describes the system configuration for simulation, this method has been tested and evaluated by using MATLAB R2021a. Operation System for this software is Windows 10 Home and its memory capacity is 6GB DDR3. Intel Core i5 @ 3.5GHz is the Matlab processor and the time required for simulation is 10.190 seconds.

In this study, tests were conducted using three different models of training and testing data: model 1 used 60% training data and 40% testing data, model 2 used 75% training data and 25% testing data, and model 3 used 80% distribution of training data and 20% testing data. Grayscale photos are used for pre-processing in the early stages of the experiment, and contrast is enhanced using MHN and filter operations. The results of segmenting images from pre-processing image input are shown in Figs. 6, 7, 8 and 9.

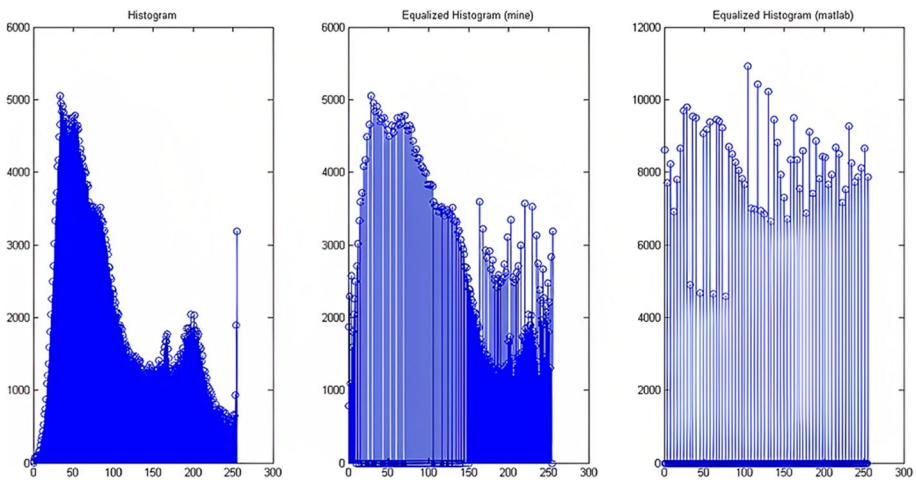
### 5.1 Chest CT-scan images dataset

The Chest CT-Scan images dataset is a collection of medical images acquired through computed tomography (CT) scanning of the chest area. This dataset is specifically curated for use in medical research and applications related to diagnosing and studying various thoracic conditions and diseases. Images are not in dcm format, the images are in jpg or png to fit the model. Data contain 3 chest cancer types which are Adenocarcinoma, Large cell carcinoma, Squamous cell carcinoma, and 1 folder for the normal cell. The images in the dataset are obtained using CT scanning, a non-invasive medical imaging technique that captures cross-sectional images of the chest region. The dataset may include images from a diverse patient population, comprising individuals with different ages, genders, and medical histories. It could cover both healthy individuals and patients with various thoracic conditions. The images cover the entire chest area, allowing visualization of various thoracic structures in a single scan. It aids in the detection and diagnosis of various thoracic conditions and diseases, such as lung cancer, pulmonary embolism, pneumonia, and interstitial lung diseases. The dataset can be used to develop and validate algorithms for automated image segmentation, object detection, and disease classification in chest CT scans.

Figures 2 and 4 typically includes multiple images arranged in a grid or a sequence, each representing different chest CT-scan slices captured from various patients. These images illustrate the variety and diversity of the dataset, depicting different anatomical structures, tissue densities, and potential abnormalities or diseases that may be present in the chest region. It allows to visually understand the characteristics of the chest CT-scan images and gain insights into the challenges associated with the automated classification of these medical images.



**Fig. 4** Sample Images from the Dataset



**Fig. 5** Histogram Normalization Result of Images

The figure showcases how the intensity distribution of pixel values in the original images is adjusted through histogram normalization. This process aims to spread out the pixel values across a wider range, ensuring that the entire intensity spectrum is fully utilized. As a result, the enhanced images exhibit improved contrast, making it easier for subsequent image analysis and classification algorithms to identify and differentiate different anatomical structures or potential abnormalities.

Figure 6 shows that the dice coefficient (DC) training and validation scores are favourable and close to one another, indicating that the MDSS-Unet model does not overfit or underfit the training data and hence produces better segmentation masks.

The training and validation curves are displayed in Fig. 7. Due to the small input image size and low resolution, MDSS U-Net trains efficiently. Up to 160 epochs, the model performs better before overfitting sets in, and with high-performance GPUs, better results would be obtained with increased image quality and batch size. When the suggested model is trained from scratch, the training process should take longer on low-performance GPUs to achieve a similar level of accuracy.

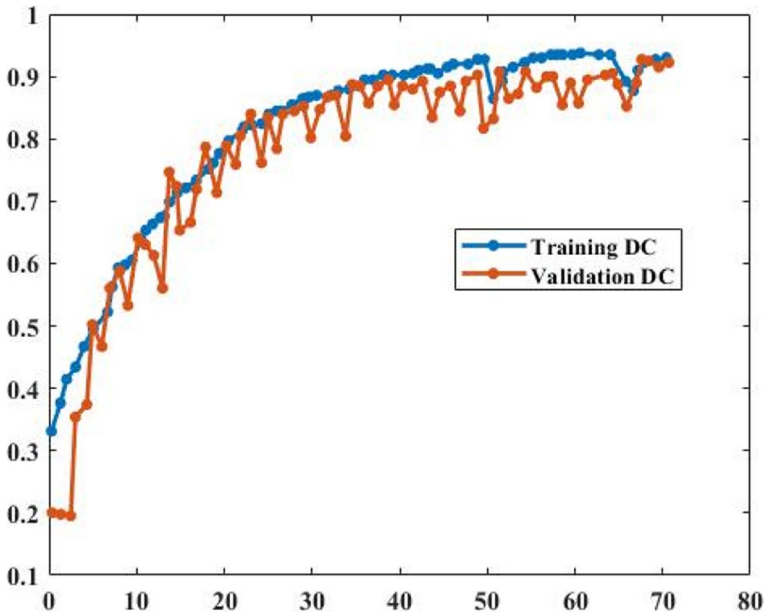


Fig. 6 Dice Coefficient Training and Validation

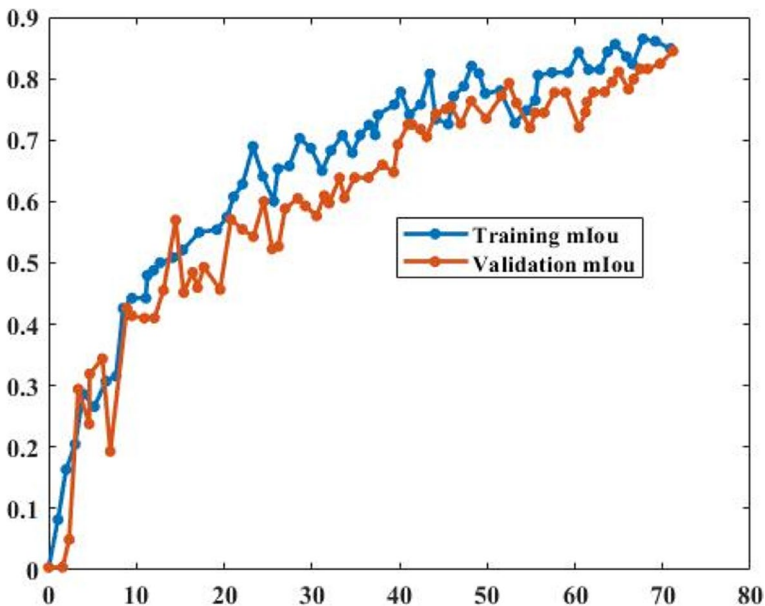


Fig. 7 Performance Analysis of MIOU

The precision-recall curves for the test sets in the MDSS U-Net architecture are shown in Fig. 8. Figure 8a displays an average recall training for the U-Net of 0.92 and a validation recall of 0.98. The average precision for training the MDSS U-Net is 0.67, and for the validation procedure, it is 0.88, representing a relative improvement of 31%.

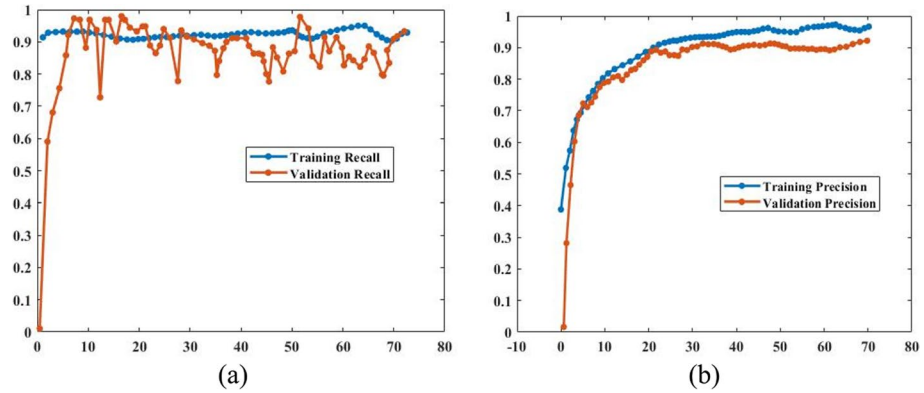


Fig. 8 Precision and Recall Training and Validation of SS U-Net

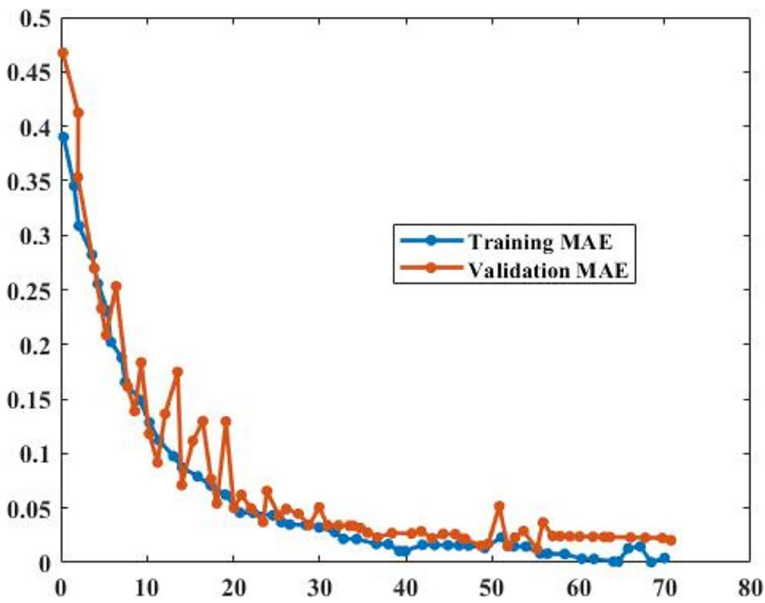


Fig. 9 Evolution of Training and Validation MAE

The training and testing model for MAE in the MDSSU U-net model is shown in Figure. Again, using the MAE, the model’s performance for such input was assessed; it was, respectively, 0.37 for training and 0.46 for validation.

Figure 10 demonstrates the selection of ideal values using the accuracy analysis of the EHO algorithm. The best values for hyper-parameters  $\alpha$  and  $\beta$  are discovered using the EHO approach, as illustrated in Figure, to reach the highest accuracy of 95%. When direct intensity measurements are utilised to initialise the position of the elephants, the EHO transform-based classifier with swarm achieves a classification accuracy of 95.2.

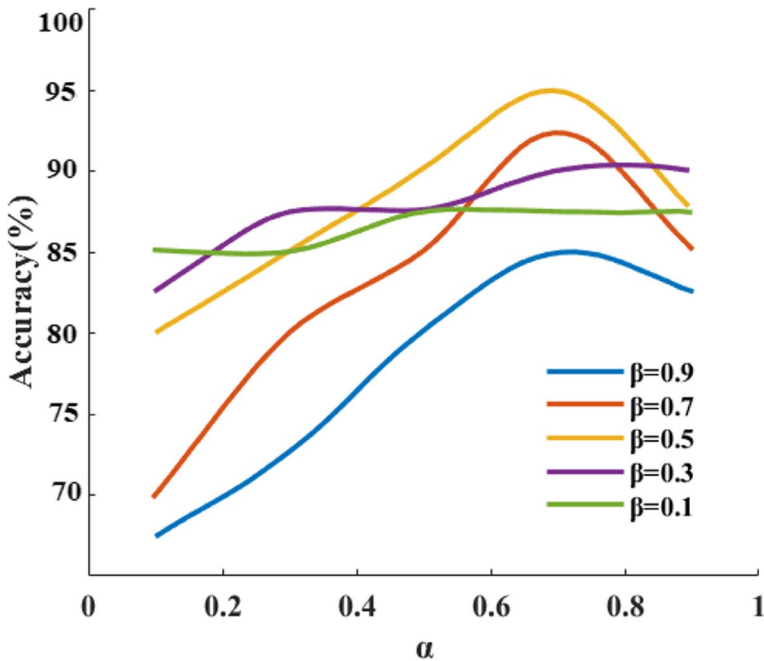


Fig. 10 Optimal Values Accuracy Analysis

## 5.2 Performance evaluation of the proposed method

Accuracy, sensitivity, and specificity are factors considered in the calculations used to determine the experiment's outcomes. Testing outcomes have been displayed using a confusion matrix, as seen in Fig. 11. A confusion matrix is a kind of comparison table that contrasts the outcomes obtained experimentally with those anticipated by CNN models. Accuracy, precision, recall (or sensitivity), specificity, Dice, and ROC or AUC are the performance metrics calculated. Equations specify the following as the performance measures: The following is what the theorems on accuracy, precision, and recall imply:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (21)$$

$$Precision = \frac{TP}{TP + FP} \quad (22)$$

$$Recall = \frac{TP}{TP + FN} \quad (23)$$

The % age of actual positive cases that were correctly predicted is determined by sensitivity. This statistic assesses the model's capacity for prediction. The following equation can be used to determine the sensitivity.

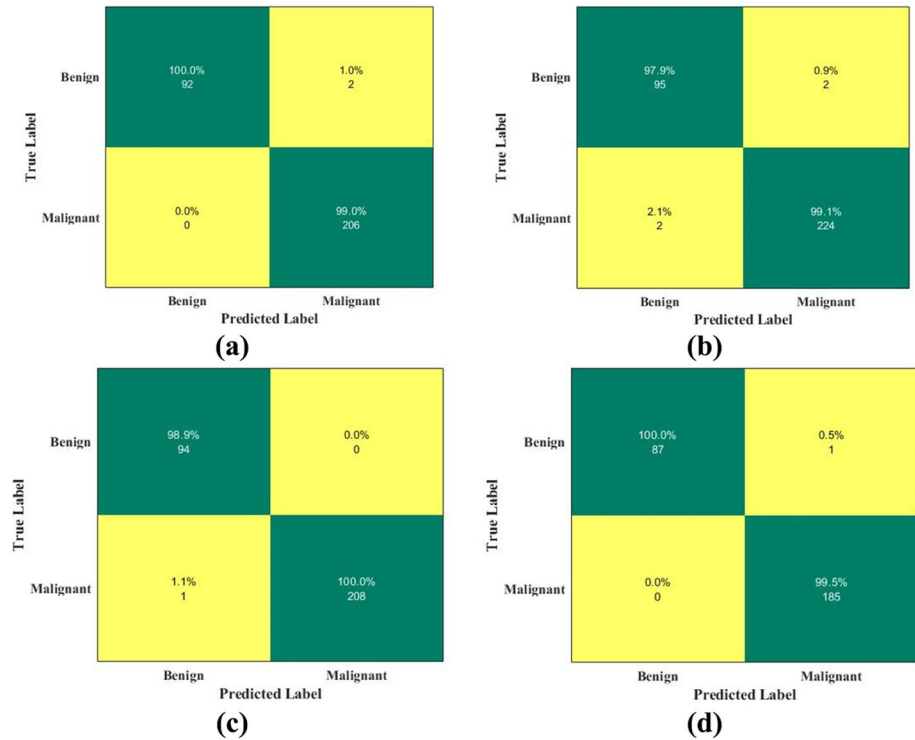


Fig. 11 C-Net Confusion Matrix for Various Factors

$$Sensitivity = \frac{TP}{TP + FN} \tag{24}$$

It was predicted accurately to employ specificity to define the %age of negative cases. A model’s ability to forecast true-negative cases of a particular category is measured using the statistic known as specificity. To interpret the outcome, these metrics were applied to each classification model. The following is the equation for calculating specificity.

$$Specificity = \frac{TN}{FP + TN} \tag{25}$$

DICE is seen as being superior because it assesses the precision of the segmentation boundaries in addition to counting the number of correctly labelled pixels. Additionally, DICE is frequently used to assess system performance repeatability via cross-validation. The segmentation is indicated by the symbol *s* in the dice coefficient.

$$DICE = \frac{2 | S_{Groundtruth} \cap S_{automated} |}{| S_{Groundtruth} | \cup | S_{automated} |} \tag{26}$$

Figure 11 displays the confusion of C-net’s matrix for varying factors. Employing several magnification factors (40X, 100X, 200X, and 400X) and numerous evaluation measures, a C-Net model was applied to the dataset. The confusion matrices make it possible to compare the models’ results in detail. The highest MCC value determined by the confusion

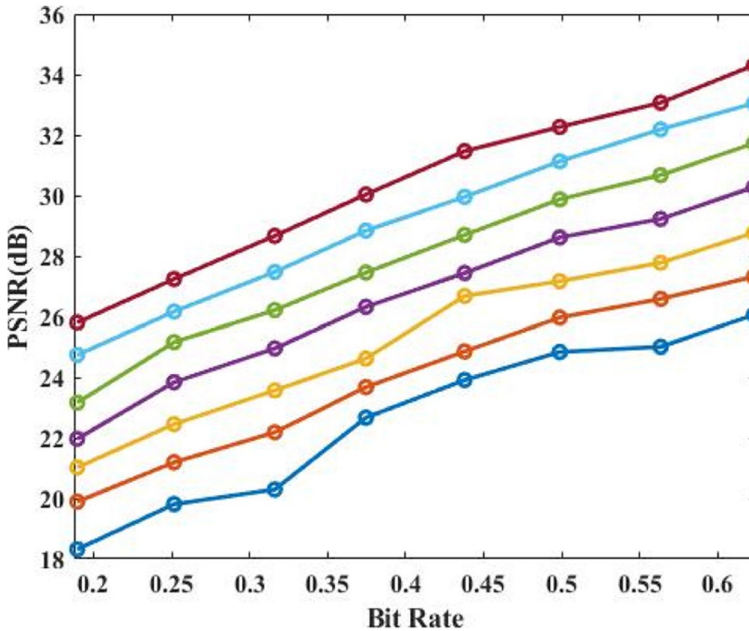


Fig. 12 Peak Signal-to-Noise Ratio

matrix, 99.23%, is 2.63% higher than the highest MCC score determined by the C-Net model, 99.23%. For these two measures, the C-Net model performs flawlessly at various magnification factors, including 40X, 100X, and 400X.

The effectiveness of the filter is compared in this article to the most widely used image compression standards, including JPEG-LS, JPEG2000, IHINT, HOP-LSC, and AVC based on the dataset. Figure 12 shows the Rate-Distortion (RD) curves for each compression method. Overall, the findings show that the performance of the suggested technique is commendable. In view of this, it is practical to use predictive coding for high-resolution image compression.

The average area under the ROC curve (AUC) for various  $\alpha$  values between 0.0 and 0.4 is compared in Fig. 13. The area will be divided into two sections by the ROC for a random classifier, and the area under this line is 0.5. The AUC value ranged between 0.882 and 0.906, which was close to one. It is clear from this that the method outperforms baseline techniques, including true positive rate and false positive rate. Thus, C-Net greatly reduces the complexity and redundancy of the network while improving the network's tolerance to spatial fluctuations that are frequently present in biological images.

By analysing 100 images of the test data, computing and visualising the segmentation metrics as a function of the threshold value, and performing a quantitative evaluation of the performance of the crack segmentation by application of a threshold to the dense dilated feature map. Figure 14 presents the outcomes. As the threshold drops, more pixels are categorised as cracks for all subset sizes, which correlates to a large SE in terms of pixels in the three analyses of dice coefficient (12(a)), sensitivity (12(b)), and precision (12(c)).



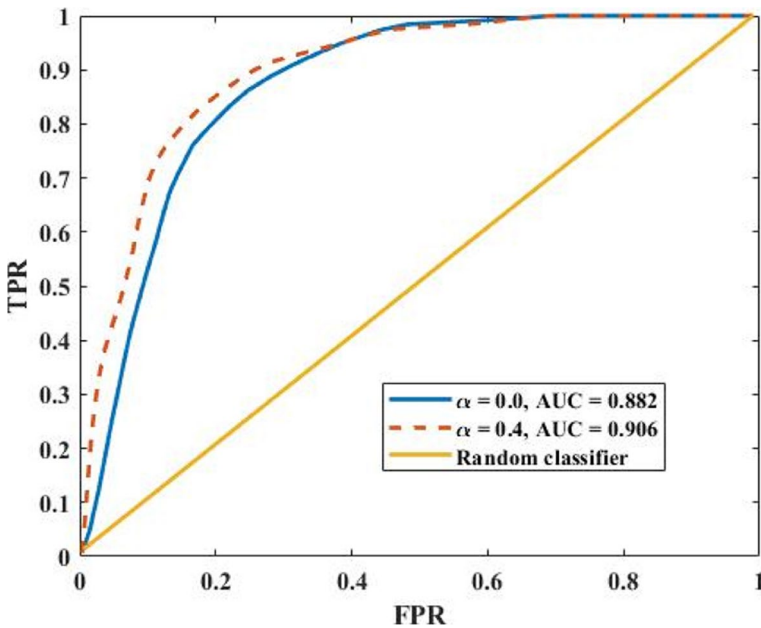


Fig. 13 AUC Curve Analysis

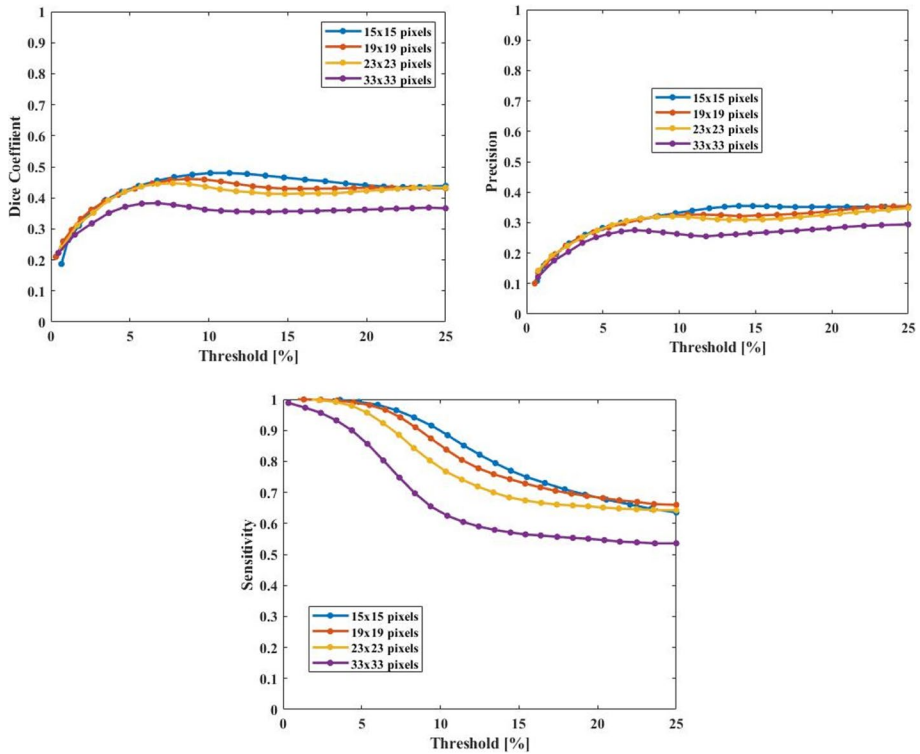
### 5.3 Comparison analysis of the proposed technique

The performance assessment of the suggested C-Net optimised hyperparameters based on the EVO algorithm is covered in this section. Comparisons have been made between the proposed hyperparameter-tailored CNN and various ANN- and deep learning-based classifiers, including ELM, CNN, MLP, and DNN. For MRI image datasets, various performance metrics based on classification accuracy have been noted. In comparison to other approaches taken into consideration for testing and comparison, the outcome of hyperparameter values that minimise the cost as produced by employing the proposed biomedical image classifier is exhibiting good results.

Figure 15 presents the comparison results between the suggested tuned model and the overall model performed on the Glioma dataset using different techniques. To compare positive performance metrics like accuracy, precision, recall, specificity, and F-score, quality criteria are considered. It is evident that CNN has hyperparameter tweaked beats, and other models, in terms of quality metrics and error rates.

The quantitative evaluation criteria for the segmentation outcomes of the various network models on these five test samples with epochs are shown in Fig. 16 along with the findings. Seg-Net has the lowest Dice coefficient of 0.912, while the highest is 0.960, as can be seen in the image. The evaluation indicators acquired using the suggested method have higher coefficients and are notably superior to those obtained using the other four network models.

The model plots the average computational time versus the total number of test cases in the dataset to better understand the scalability of C-Net EHO. Figure 17 displays the



**Fig. 14** Segmentation Metrics Threshold Analysis with Pixel

findings, with the number of tasks shown on the x-axis and the average computation time in seconds indicated on the y-axis. Keep in mind that the same number of jobs appears in numerous occurrences (x-axis value). To indicate the average computation time for the appropriate number of tasks, the average of the y-axis values of these instances is computed in this example. The figure makes it clear that the scalability greatly increases as the image feature grows. The average processing time of C-Net EHO is comparable to that of other algorithms when the number of tasks is no more than 1.

The table presents a comparison of different techniques used for biomedical image classification, along with their corresponding accuracy percentages. Each technique represents a different approach or model used to classify biomedical images. The accuracy percentage indicates the performance of each technique in correctly classifying the images. The “Proposed” technique refers to the approach presented in the current study (the one being discussed in the paper). This technique employs the multi-scale dense dilated semi-supervised U-Net (MDSSU-Net) with CNN architecture for image classification. It outperformed all other techniques, achieving the highest accuracy of 99.35%. The table demonstrates the comparative performance of various image classification techniques on the same dataset. The “Proposed” technique stands out with the highest accuracy percentage, indicating that the proposed model is the most effective among all the techniques considered in this evaluation. This validates the superiority of the proposed method for automating biomedical image classification and highlights its potential for real-world biomedical applications Fig. 18.

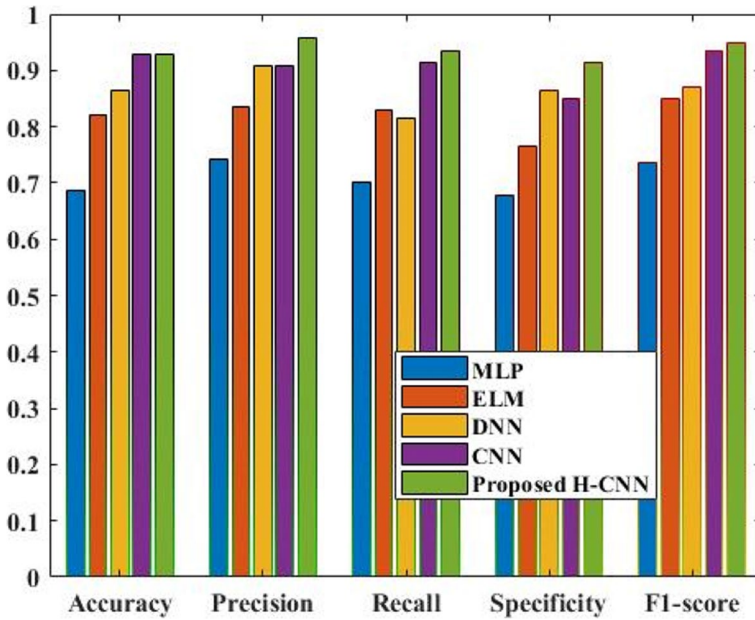


Fig. 15 Comparison Analysis of Quality Factors with ANN Techniques

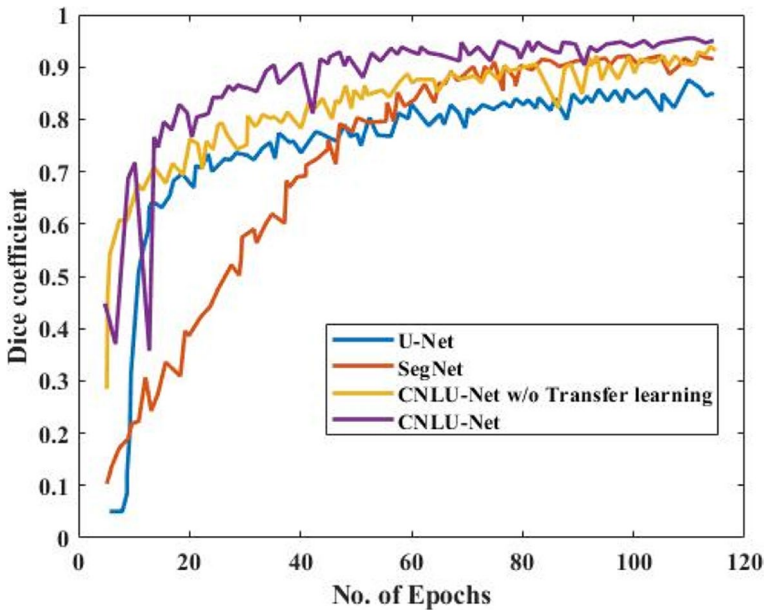


Fig. 16 Dice Coefficient Comparative Analysis

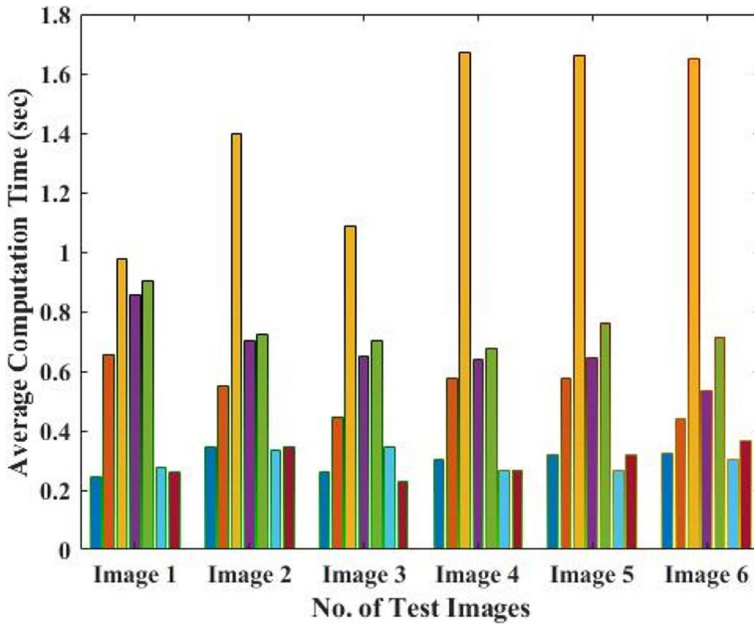


Fig. 17 Average Computation Time of Different Test Images

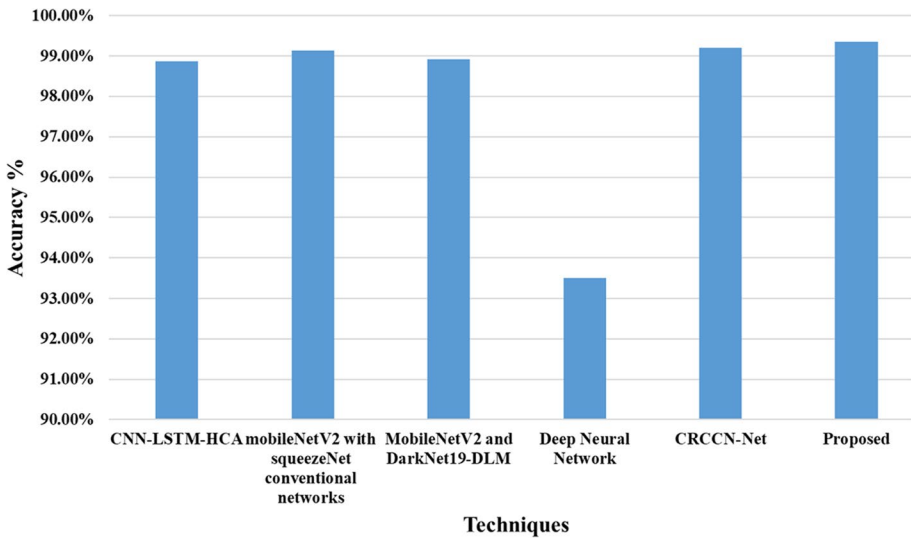


Fig. 18 Accuracy Comparison with state of art techniques

## 6 Discussions

The research presents an extensive evaluation of the proposed MDSS-Unet model for segmentation tasks using various performance metrics and comparative analyses. The dice coefficient

(DC) training and validation scores are observed to be favourable and close to one another. This indicates that the MDSS-Unet model does not suffer from overfitting or underfitting, resulting in better segmentation masks. The MDSS U-Net exhibits efficient training, attributed to the small input image size and low resolution. However, there is a slight performance drop due to overfitting after 160 epochs. Higher image quality and batch size, with powerful GPUs, are expected to yield better results. The MDSS U-Net demonstrates improved recall and precision during validation compared to training, indicating a relative improvement of 31%. The model excels in correctly classifying positive cases during both training and validation. The Mean Absolute Error (MAE) analysis shows that the MDSSU U-net model achieves a low error rate during training (0.37) and validation (0.46), suggesting good accuracy in image segmentation. The EHO algorithm effectively identifies the best values for hyperparameters  $\alpha$  and  $\beta$ , resulting in a classification accuracy of 95.2%. The proposed model shows promising results in terms of accuracy analysis. The C-Net model performs exceptionally well across various magnification factors, with an MCC value of 99.23%. It outperforms other tested models, including the true positive rate and false positive rate, demonstrating the effectiveness of the proposed technique. The suggested image compression technique proves to be commendable compared to standard compression methods, including JPEG-LS, JPEG2000, IHINT, HOP-LSC, and AVC, based on the dataset. This validates the practicality of using predictive coding for high-resolution image compression. The proposed method outperforms baseline techniques, including true positive rate and false positive rate, as evidenced by the area under the ROC curve (AUC) values ranging between 0.882 and 0.906. This demonstrates the model's improved tolerance to spatial fluctuations in biological images. Analyzing segmentation metrics as a function of the threshold value shows that reducing the threshold increases the number of pixels classified as cracks, resulting in higher sensitivity and reduced precision. The suggested hyperparameter-tuned CNN model outperforms other ANN-based classifiers, including ELM, MLP, and DNN, in terms of quality metrics and error rates. Comparing segmentation outcomes, the suggested method using MDSS-Unet shows superior performance with higher dice coefficients compared to Seg-Net and other network models. The scalability of the C-Net EHO approach improves with larger image features, and the average computation time is comparable to other algorithms.

In conclusion, the proposed MDSS-Unet model demonstrates promising results in image segmentation tasks, offering better accuracy and performance compared to other existing approaches. The research outcome validates the efficacy of the proposed method and its potential for real-world applications in various image-processing tasks.

## 6.1 Limitations

Despite the promising results, this research has certain limitations that should be acknowledged. The study may have been conducted using a limited dataset, which may restrict the generalizability of the results to a broader population. The experiments were performed on a specific MATLAB version with hardware limitations. The proposed method's performance on other systems with different configurations may vary. The performance of the proposed method may be sensitive to certain hyperparameters, necessitating further investigations for optimal parameter selection. The study utilized grayscale photos for pre-processing, and the effectiveness of the proposed method may vary with other types of image datasets. The results and conclusions drawn from this research may be specific to the chosen model and may not be directly applicable to other image classification or segmentation problems.

## 7 Research Conclusion

This study, proposed an automated biomedical image classification system using the Multi-Scale Dense Dilated Semi-Supervised U-Net with CNN Architecture (MDSSU-Net). Our system aims to address the challenges associated with accurate and efficient biomedical image segmentation, particularly in chest CT-scan images. Through extensive experimentation and evaluation, we have demonstrated the effectiveness and superiority of our proposed model. The necessity of automated biomedical image classification has become increasingly evident as the healthcare system relies heavily on medical imaging for non-invasive diagnostic treatments. The manual segmentation of biomedical images is laborious, time-consuming, and costly, especially given the growing quantity and variety of medical images. An automated computer-aided diagnostic system, like the one proposed in this study, has the potential to transform clinical operations, improve patient care, and ease the burden on medical experts.

The MDSSU-Net architecture incorporates multi-scale dense dilated residual blocks and skips connection paths, enhancing the learning capabilities and improving segmentation accuracy. The integration of semi-supervised learning further reduces the need for labelled training data, making the model more efficient and robust. Extensive experimentation on the Chest CT-Scan images dataset showcased the superiority of our proposed model compared to existing methods. The high Dice Coefficient, Mean Intersection Over Union (MIOU), recall, and precision scores indicate the model's ability to accurately and consistently classify biomedical features in CT-scan images. The model's training process is efficient, benefiting from the use of high-performance GPUs, enabling better results with increased image quality and batch size. The Elephant Herding Optimization (EHO) method employed for hyperparameter tuning contributes to achieving the highest accuracy with optimal values. The proposed MDSSU-Net architecture demonstrates strong generalization capability across different magnification factors and evaluation measures. It effectively handles unbalanced features and successfully extracts novel image attributes. The successful development of an automated biomedical image classification system has significant clinical implications. It has the potential to aid radiologists and medical professionals in diagnosing and understanding various thoracic conditions and diseases, leading to improved patient outcomes and medical decision-making.

## 8 Conclusion in Light of Experimental Results

Based on the experimental results and evaluation metrics, our proposed automated biomedical image classification system, MDSSU-Net, shows promising potential for practical implementation in real-world clinical settings. It addresses the technical gaps associated with accurate and efficient automated segmentation, supporting the medical community in diagnosing thoracic diseases with high precision. The findings from this study have contributed significantly to the field of medical imaging and computer-aided diagnosis. Our work demonstrates the feasibility of using advanced deep learning techniques for automated biomedical image classification, emphasizing the importance of leveraging multi-scale dense dilated architectures and semi-supervised learning to achieve superior results. The success of the MDSSU-Net model in accurately classifying biomedical features in chest CT-scan images opens up new avenues for future research and applications. As the field of medical imaging continues to evolve, our proposed approach can serve as

a foundation for developing even more sophisticated and efficient automated systems that enhance patient care and medical diagnostics. The study firmly believes that the insights and contributions from this study will positively impact the healthcare industry, enabling faster and more accurate disease diagnosis, and ultimately improving the well-being of patients worldwide.

## 9 Future work

While the proposed MDSSU-Net model has shown excellent performance, there is still room for further improvement and exploration. In the future, we plan to focus on the following aspects to enhance the capabilities and applicability of our automated biomedical image classification system:

Integrating information from multiple imaging modalities, such as MRI or X-ray, can provide a more comprehensive and accurate assessment of thoracic conditions. Investigating the application of transfer learning techniques to fine-tune the MDSSU-Net model on specific disease-related datasets can potentially improve performance and reduce the need for large labelled datasets. Enhancing the explainability and interpretability of the MDSSU-Net model can facilitate trust and adoption by medical professionals, allowing them to understand and validate the model's decisions. Optimizing the model for real-time implementation on low-powered devices can enable point-of-care diagnosis and enhance accessibility in resource-limited settings. By addressing these aspects, the study aims to advance the field of automated biomedical image classification and contribute to the broader adoption of computer-aided diagnosis in medical practice. This work lays the groundwork for future research and applications that have the potential to revolutionize healthcare and improve patient outcomes worldwide.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11042-023-16659-1>.

**Data availability** Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

## Declarations

**Conflict of interest** The authors declare no conflict of interest.

## REFERENCES

1. Haque IRI, Neubert J (2020) Deep learning approaches to biomedical image segmentation. *Inform Med Unlocked* 18:100297
2. Kist AM, Döllinger M (2020) Efficient biomedical image segmentation on EdgeTPUs at point of care. *IEEE Access* 8:139356–139366
3. Hu X, Yang H (2020) DRU-net: a novel U-net for biomedical image segmentation. *IET Image Process* 14(1):192–200
4. Weng W, Zhu X (2021) INet: convolutional networks for biomedical image segmentation. *IEEE Access* 9:16591–16603
5. Tchito Tchagpa C, Mih TA, Tchagna Kouanou A, Fozin Fonzin T, Kuetche Fogang P, Mezatio BA, Tchiosop D, 2021. Biomedical image classification in a big data architecture using machine-learning algorithms. *J Healthcare Eng*, 2021
6. Srivastava A, Jha D, Chanda S, Pal U, Johansen HD, Johansen D, Riegler MA, Ali S, Halvorsen P (2021) Msrf-net: A multi-scale residual fusion network for biomedical image segmentation. *IEEE J Biomed Health Inform* 26(5):2252–2263

7. Isensee F, Jaeger PF, Kohl SA, Petersen J, Maier-Hein KH (2021) NnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods* 18(2):203–211
8. Punn NS, Agarwal S (2022) Modality specific U-Net variants for biomedical image segmentation: a survey. *Artif Intell Rev*:1–45
9. Valanarasu JMJ, Sindagi VA, Hacihaliloglu I, Patel VM (2021) Kiu-net: Overcomplete convolutional architectures for biomedical image and volumetric segmentation. *IEEE Trans Med Imaging* 41(4):965–976
10. Gudhe NR, Behravan H, Sudah M, Okuma H, Vanninen R, Kosma VM, Mannermaa A (2021) Multi-level dilated residual network for biomedical image segmentation. *Sci Rep* 11(1):1–18
11. Assad MB, Kiczales R (2020) Deep biomedical image classification using diagonal bilinear interpolation and residual network. *Int J Intel Netw* 1:148–156
12. Gong P, Yu W, Sun Q, Zhao R, Hu J (2021) Unsupervised Domain Adaptation Network with Category-Centric Prototype Aligner for Biomedical Image Segmentation. *IEEE Access* 9:36500–36511
13. Xiao F, Shen C, Chen Y, Yang T, Chen S, Liao Z, Tang J, 2021. RCGA-Net: An Improved Multi-hybrid Attention Mechanism Network in Biomedical Image Segmentation. In 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (pp. 1112–1118). IEEE.
14. Cao B, Tu S, Xu L, 2021. Flexible-CLmsr: Regularized Feedback Connections for Biomedical Image Segmentation. In: 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (pp. 829–835). IEEE
15. Sarwinda D, Paradisa RH, Bustamam A, Anggia P (2021) Deep learning in image classification using residual network (ResNet) variants for detection of colorectal cancer. *Procedia Comput Sci* 179:423–431
16. Das A, Mohapatra SK, Mohanty MN (2022) Design of deep ensemble classifier with fuzzy decision method for biomedical image classification. *Appl Soft Comput* 115:108178
17. Alnabhan M, Habboush AK, Abu Al-Haija Q, Mohanty AK, Pattanaik S, Pattanayak BK, 2022. Hyper-Tuned CNN Using EVO Technique for Efficient Biomedical Image Classification Mobile Information Systems, 2022.
18. Barzekar H, Yu Z (2022) C-Net: A reliable convolutional neural network for biomedical image classification. *Expert Syst Appl* 187:116003
19. Orazayeva A, Tussupov J, Wójcik W, Pavlov S, Abdikerimova G, Savytska L (2022) Methods for Detecting and Selecting Areas on Texture Biomedical Images of Breast Cancer. *Informatyka, Automatyka, Pomiar w Gospodarce i Ochronie Środowiska* 12(2):69–72
20. Sagar A, 2022. Uncertainty quantification using variational inference for biomedical image segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (pp. 44–51)
21. Ma T, Dalca AV, Sabuncu MR, 2022. Hyper-convolution networks for biomedical image segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (pp. 1933–1942)
22. Drees D, Eilers F, Jiang X, 2022. Hierarchical Random Walker Segmentation for Large Volumetric Biomedical Images. *IEEE Trans Image Process*
23. Olimov B, Sanjar K, Din S, Ahmad A, Paul A, Kim J (2021) FU-Net: fast biomedical image segmentation model based on bottleneck convolution layers. *Multimedia Systems* 27(4):637–650
24. Alanazi AA, Khayyat MM, Khayyat MM, Elamin Elnaim BM, Abdel-Khalek S, 2022. Intelligent Deep Learning Enabled Oral Squamous Cell Carcinoma Detection and Classification Using Biomedical Images Computational Intelligence and Neuroscience, 2022
25. Hamza MA, Albraikan AA, Alzahrani JS, Dhahbi S, Al-Turaiki I, Al Duhayyim M, Yaseen I, Eldesouki MI, 2022. Optimal Deep Transfer Learning-Based Human-Centric Biomedical Diagnosis for Acute Lymphoblastic Leukemia Detection Computational Intelligence and Neuroscience, 2022.
26. Behera M, Sarangi A, Mishra D, Mallick PK, Shafi J, Srinivasu PN, Ijaz MF (2022) Automatic data clustering by hybrid enhanced firefly and particle swarm optimization algorithms. *Mathematics* 10(19):3532
27. Swamy SR, Praveen SP, Ahmed S, Srinivasu PN, Alhumam A (2023) Multi-features disease analysis based smart diagnosis for covid-19. *Comput Syst Sci Eng* 45(1):869–886
28. Yang J, Shi R, Wei D, Liu Z, Zhao L, Ke B, Pfister H, Ni B (2023) MedMNIST v2-A large-scale lightweight benchmark for 2D and 3D biomedical image classification. *Scien Data* 10(1):41
29. Pradhan AK, Das K, Mishra D, Chithaluru P (2023) Optimizing CNN-LSTM hybrid classifier using HCA for biomedical image classification. *Expert Syst* 40(5):e13235
30. Nazir S, Dickson DM, Akram, M.U., 2023. Survey of explainable artificial intelligence techniques for biomedical imaging with deep neural networks. *Comput Biol Med*, p.106668.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted



manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

## Authors and Affiliations

**Anoop V<sup>1</sup> · Bipin PR<sup>2</sup> · Anoop BK<sup>3</sup>**

✉ Anoop V  
vanoop@gmail.com

Bipin PR  
bipinpr@gmail.com

Anoop BK  
dranoopbk@gmail.com

<sup>1</sup> Department of Artificial Intelligence and Data Science, Jyothi Engineering College, Cheruthuruthy, Thrissur, Kerala, India

<sup>2</sup> Department of Electronics & Communication Engineering, Adishankara Institute of Engineering and Technology, Kalady, Kerala, India

<sup>3</sup> Department of Artificial Intelligence & Machine Learning, Srinivas Institute of Technology, Mangalore, Karnataka, India