

AfroNet: A Cross-Attention Enhanced U-Net for Breast Cancer Image Segmentation

Vuppula Manohar ^{1,*}, P. S. Rao ², Sreedhar Kollem ³, Karri Chiranjeevi ^{4,*}, B. Jaya ⁵,
M. Shashidhar ⁵, Syed M. Ahamed ¹, Appala S. Kumar ¹, and Manasa Koppula ¹

¹Electronics & Communication Engineering, Vaagdevi Engineering College, Warangal, Telangana, India

²Electronics & Communication Engineering, CVR College of Engineering, Hyderabad, Telangana, India

³Electronics & Communication Engineering, School of Engineering, SR University, Warangal, Telangana, India

⁴Electronics & Communication Engineering, A.U. College of Engineering, Andhra University, Visakhapatnam, Andhra Pradesh, India

⁵Electronics & Communication Engineering, Vaagdevi College of Engineering, Warangal, Telangana, India

Email: donemanoharvu@gmail.com (V.M.); srinivasarao@cvr.ac.in (P.S.R.); ksreedhar829@gmail.com (S.K.);
chiru404@gmail.com (K.C.); jaya.bangari15@gmail.com (B.J.); sasi47004@gmail.com (M.S.);

musthak.ahmed@vecw.edu.in (S.M.A.); appala.sravan@gmail.com (A.S.K.); manasa@vecw.edu.in (M.K.)

*Corresponding author

Abstract—Accurate segmentation of medical images is essential for reliable breast cancer detection and for downstream tasks such as identifying nuclei and cell membranes in histopathology slides. However, traditional deep learning models like U-Net often encounter limitations in precision, computational efficiency, and adaptability when confronted with complex, multimodal datasets. To address these challenges, we propose AfroNet, a novel U-shaped deep learning architecture that integrates advanced attention mechanisms specifically designed for breast cancer image segmentation. AfroNet introduces three complementary modules: (1) a cross-attention module that adaptively recalibrates encoder–decoder interactions to emphasize diagnostically relevant semantic features; (2) a multi-scale feature-fusion block that captures fine spatial details across varying resolutions to enhance boundary delineation; and (3) an adaptive skip-connection enhancement strategy that strengthens gradient flow and preserves contextual information throughout the network. Extensive experiments on benchmark breast cancer histopathology datasets demonstrate that AfroNet consistently outperforms state-of-the-art segmentation methods in terms of Dice Coefficient (DC), Intersection-over-Union (IOU), and inference speed. These results highlight AfroNet’s potential as a robust and efficient framework for high-precision breast cancer histopathology analysis and clinical decision support.

Keywords—breast cancer, semantic segmentation, U-Net, breast image, pyramidal network

I. INTRODUCTION

Breast cancer is a common and possibly fatal illness that affects many people worldwide. This occurs when aberrant cell growth in breast tissue leads to malignancies. According to data on global health, breast cancer is a widespread disease diagnosed in women. It is the root of

numerous cancer-related fatalities globally [1]. The high death rates connected with breast cancer highlight the necessity for efficient diagnosis and detection techniques. Improving the health of patients and lowering death rates depend heavily on early detection. Therefore, medical research must develop precise and effective approaches for identifying breast cancer. The healthcare system is an essential part of society because it ensures that every person receives the appropriate analysis and treatment. It also lowers the need for trained labour and results in improved care and prescription medicines. In today’s fast-paced and technologically advanced world, integrating new technology into healthcare has become essential and unavoidable. Thus, a crucial first step in the advancement of medical research is the integration of various technologies into healthcare.

In recent times, Machine Learning (ML) and Deep Learning (DL) have been more effective in identifying breast cancer. DL uses multilayered artificial neural networks to learn intricate ordered data illustrations. A key idea in DL is Transfer Learning (TL), which transforms the field using models that have been previously trained and learned on huge datasets for tasks. These pre-trained models are significant resources for feature extraction in numerous areas, such as breast cancer diagnosis, because they learn complex and relevant features from large amounts of data. By utilizing the information included in these trained models, we can accelerate the training procedure, improve model efficiency, and overthrow constraints imposed by inadequate data.

However, ML uses a variety of algorithms to interpret and analyse data, find patterns, and develop predictions. With the use of fresh, untainted data, these algorithms can recognize patterns in labelled datasets and provide accurate predictions. There are several advantages of

using DL and ML algorithms for breast cancer screening. These methods can provide a strong foundation to boost patient care, help medical professionals make wise judgments, and increase the precision and efficiency of diagnostic procedures.

Feature extraction is an essential step in the study of images of breast cancer. More casually depicting complicated data requires locating and extracting discriminative structures and patterns from medical images. The buried information inside the photos can be uncovered by extracting pertinent elements such as statistical measurements, texture features, or shape qualities. This procedure is crucial because it makes it possible to identify image-based biomarkers, characterize tumors more easily, and distinguish between healthy and malignant cells. Enhancing precision, effectiveness, and dependability of breast cancer detection and treatment planning requires an accurate feature extraction. This study presents the U-Net model, a one-level method that concurrently detects, segments, and categorizes aberrant masses in mammograms as both benign and malignant. Using this method, every pixel in a mammography image is categorized as benign, malignant, or normal. Its capacity to function without preprocessing or weights trained in advance is another benefit. To reduce manual intervention and improve algorithm generalization, this method uses a deep learning architecture that combines Convolutional Neural Network (CNN) and Transformer structures with morphological structural constraints to design a three-dimensional segmentation of medical image models for multi-objective joint detection. Specifically, UMP-UNet (using a U-shaped network to alter morphological constraints) is the suggested multi-objective segmentation network model depending on combined network learning and morphologic structure constrictions. The reference network in the proposed model is the U-Net architecture Trans-U-Net, that is built collaboratively using CNN and Transformer structures. A previous learning method of the morphological structure was introduced based on the reference network to help with restrictions to further enhance the overall efficacy of multi-objective joint segmentation of 3D medical images, and two publicly accessible datasets were used for research and analysis. The findings of these studies reveal the efficacy of proposed method for 3D medical image segmentation.

Breast cancer remains one of the most prevalent and life-threatening conditions among women globally. Accurate and automated segmentation of breast cancer regions in histopathological or radiological images is vital for diagnosis, treatment planning, and outcome monitoring. However, existing segmentation models often struggle with inconsistent boundaries, low contrast, and variability in tumor size and shape. These challenges motivate the development of more robust and context-aware architectures, such as the proposed AfroNet, to enhance segmentation accuracy and clinical reliability.

The key contributions of this paper are as follows:

- We propose AfroNet, a novel attention-enhanced U-shaped architecture tailored for breast cancer image segmentation.

- We introduce a cross-attention module that strengthens encoder-decoder interactions by selectively attending to informative spatial features.
- A multi-scale feature fusion strategy is incorporated to capture rich contextual information at different resolutions.
- An adaptive skip connection enhancement is designed to preserve semantic integrity and improve gradient flow.
- Extensive experiments on benchmark datasets show that AfroNet outperforms state-of-the-art models in both accuracy and robustness.

The rest of this study is organized as follows. Section II explains current approaches for classifying and segmenting breast masses that have been proposed over the last ten years. Section III provides an overview of the proposed scheme. The datasets used to train and test the suggested models are described in Section IV, as are the metrics used to evaluate the effectiveness of the suggested method. Section V provides examples of the results of semantic segmentation of breast masses. The advantages and disadvantages of our suggested approach are discussed in Section VI, along with our plans for future research.

II. LITERATURE REVIEW

Elevated breast density on mammograms is a significant risk factor for breast cancer. Full-Field Digital-Mammography (FFDM) requires automated, precise, and repeatable breast density assessment to enable clinical applications. Using FFDM data, we evaluated a novel automated breast density measurement Percent Dense area (PDA) and compared it with the standard operator-assisted method (PD). Conditional logistic regression was used to estimate the Odds Ratios (ORs) for breast cancer, with adjustments made for body mass index. All measurements had 95% Confidence Intervals (CI) attached to them. According to Sparse Denoising Network (SDN) analysis, PDA requires a nonlinear connection amid the mammographic signal and its fluctuation or a breast density biomarker [2]. Although the annual death rate is rising, patients' chances of survival will significantly increase if breast cancer tumors are identified sooner and treated appropriately immediately. The dataset was subjected to both Support Vector Machines (SVM) and Artificial Neural Networks (ANN) to categorize breast cancer tumors and compare their respective performance. The SVM employing radial functions obtained highest identification accuracy of 91.6%, whereas the ANN reached 76.6%. Consequently, an SVM was used to determine pertinent risk variables for breast cancer [3].

Research has concentrated on investigating bereavement following a cancer diagnosis, bereavement among parents of a child lost to cancer and bereavement among caregivers of patients with cancer. The study used a qualitative case study methodology, with ten breast cancer patients ranging in age from 47 to 54 years old. Thematic analysis revealed that the mental health of breast cancer patients was adversely influenced by the loss of

companions. Patients with breast cancer who are grieving can experience a wide range of symptoms, including depression, intense anxiety, difficulty sleeping, loneliness, and isolation. The findings also revealed several risk variables, including social relationships, remorse, self-blame, intense emotional commitment to a spouse, isolation during the COVID-19 pandemic, and guilt [4]. The performance and scalability constraints of current approaches highlight the need for further investigation.

Using a pre-trained ResNet50V2 model, we offer a hybrid reliable breast cancer detection strategy that combines the capability of DL with ensemble-based machine learning techniques. The results of our thorough testing offer strong proof of the resilience and excellent performance of proposed approach. In comparison to state-of-the-art models, our approach attained a higher accuracy of 95%, along with a precision 94.86%, recall 94.32%, and F1-Score 94.57% [1].

DL techniques have the latent to advance breast cancer cell identification, reduce false positives, and shorten the time required for human breast cancer diagnosis. To analyse the biopsy survival of patients with breast cancer, this research examined the accuracy of ANN, constrained Boltzmann machines, deep-autoencoders, and CNN. With an accuracy score of 0.97, the Deep Autoencoders achieved the second highest accuracy score after the Restricted Boltzmann Machine. ANN received the lowest accuracy of 0.89, whereas CNN attained 92% accuracy [5]. Owing to their intricate processing, most approaches require a high degree of computational complexity. The study recommends using feature extraction and optimization to diagnose breast cancer. We suggest using Neural Networks (NNs) as the basis for classifiers to identify cancer.

The Curated Breast Imaging Subset of the Digital Database for Screening Mammography (CBIS-DDSM) cancer image dataset was used to evaluate this approach. Compared to existing breast cancer algorithms, such as CNN, Ransom-Random Forest (RF), SVM, and NN, the proposed technique has demonstrated efficacy [6]. A U-Net-based efficient segmentation technique for extracting Regions of Interests (Rols) from mammograms is presented. The goal is to increase the classification accuracy by fusing a Case-Based Reasoning (CBR) approach with DL. While CBR bids a clear and accurate classification, DL extends precise mammography segmentation. Testing on the dataset CBIS-DDSM, the proposed method outperformed several renowned ML and DL techniques, achieving outstanding results with an accuracy of 86.71% and recall of 91.34% [7]. New methods for evaluating medical images are labour-intensive, costly, time-consuming, and error prone. It would be advantageous to use a computer-aided tool that can automatically make diagnosis and treatment decisions. A high-resolution multi-view deep-CNN with fuzzy-based visual analysis is suggested for detection of breast cancer by means of an SVM classifier. Results gathered from DDSM Curated Breast Imaging Subset, and the Mammography Screening Information Database indicate that the VGG is more dependable and thousands of times

faster than previous programmed anatomy segmentation [8]. As ultrasound imaging is inexpensive and less intrusive, it is a routine diagnostic method for this type of condition.

However, because this method is subject to some level of uncertainty, computer-supported methods have recently been suggested to lessen operator workload and increase diagnostic efficiency. A complete pipeline for automated deep learning was suggested for the segmentation and categorization of breast ultrasound images. The results of trials directed on publicly available datasets validate effectiveness of ensemble approaches compared with individual networks. Moreover, our optimal configuration demonstrates competitiveness against the state-of-the-art, achieving 91% accuracy in the classification task and 82% DC in the segmentation test [9]. Research based on machine learning that makes great use of naïve bays, KNN, SVM, and decision trees is one example of successful research in this area. These discoveries have resulted in the creation of improved algorithms.

Conversely, a relatively new method called DL has been accustomed to classifying breast cancer. The algorithms for deep learning performed better than those for machine learning. The photos' most interesting parts were cropped. CNNs are frequently used by academics to classify photos. Briefly, CNN is the method most employed for image classification [10]. The significance of machine learning in the development of convolutional neural networks and image segmentation algorithms, which have demonstrated remarkable efficacy in image analysis tasks, has been emphasized [11]. The process of classifying images as either cancerous or non-cancerous requires many tasks, such as preprocessing images, extraction of features, classification, and analysis. These findings suggest that the most effective cancer diagnostic techniques used today are deep learning techniques [12].

ResNet18, InceptionV3, and Shuffle Net are deep neural networks used to classify breast cancer in binary histopathology images. The networks were pre-trained via transfer-learning using database ImageNet, and their output-layers were fine-tuned using histopathological images from public dataset Break His. ResNet18 scored highest accuracy of 98.73% for the 2-class classification of benign/malevolent cases, followed by Shuffle Net with 97.65% and Inception-V3Net with 97.44% [13].

For histopathological image classification, a deep-CNN-based transfer learning strategy is suggested, together with structured filter pruning, to condense the run-time resource. VGG19, ResNet34, and ResNet50, three widely used pre-trained CNNs, were used in several extensive tests. By using the VGG19 pruned model, we were able to reduce 63.46% of FLOPs and obtain 91.25% accuracy, exceeding previous approaches on same architecture and dataset. In contrast, accuracy increased to 91.80%, with 40.63% less FLOPs when ResNet34 pruned model was used. Furthermore, attained 92.07% accuracy with 30.97% fewer FLOPs by using the ResNet50 model [14]. Conventional ML/DL techniques can identify existence of a lesion & categorize as benign or malignant, that may be crucial for reducing interpretation time and

increasing accuracy. Subsequently, studies concerning the prediction of breast cancer risk using mammography, which might enable the customization of screening programs concerning both frequency and modality, were examined. Various augmentation techniques that overcome the absence of labelled data have been examined. The article gives a summary of the potential benefits of Artificial Intelligence (AI) in the field of breast imaging, highlighting both its potential and the challenges that still need to be addressed [15].

CNNs can be utilized in the framework of ultrasound image-based breast cancer forecast to evaluate and classify images of breast tissue as benign or malignant. To do this, the properties of breast tissue were compared to known normal breast tissue [16]. A model was created using a CNN-based TL to notice breast cancer from mammography images. The developed structure consists of multiple stages: data augmentation; breast region extraction; and a Gaussian filter. The experimental testing ground was a mini-MIAS dataset, with the highest accuracy of 95.71%. Compared to the existing approaches, the new framework performs noticeably better [17]. With the advent of Computer-Assisted Detection (CAD), the crucial problem in breast cancer identification is human interpretation. Artificial intelligence and several computer vision-based techniques are being developed for the diagnosis and treatment of breast cancer using machine and deep learning [18]. While there are limitations to traditional diagnostic procedures, artificial intelligence techniques, such as machine learning and deep learning, provide the possibility of a more precise and effective diagnosis. With encouraging findings, researchers have created state-of-the-art DL models to identify lymph node metastases from breast cancer images. Combining patient data and radiographic data will improve the accuracy of these models [19]. In the realm of risk analysis, ML/DL techniques yield encouraging outcomes, whether in the form of risk group classification or risk score prediction [20]. Deep learning models, even interpretable models, have significant obstacles to being widely accepted in clinical settings when they use contradicting or irrelevant data to generate predictions. To mitigate this issue, pseudo-class component prototype-networks for interpretable breast cancer classification have been suggested. The suggested strategy successfully increased accuracy and interpretability by 8% and 18%, respectively, according to the experimental findings on the Break His dataset [21]. Although still early in its development, AI is already being used to educate patients and assist physicians in their work. Future studies should investigate how AI-powered resources can improve knowledge and help make better decisions about breast cancer screening, particularly for low-literacy and vulnerable populations [22]. UCFilTransNet is a lightweight transformer-based segmentation model that addresses the semantic gap in U-Net by introducing a Cross-Filter Transformer (CFTrans) block and a Residual Pyramid Squeeze-Excitation (RPSE) module. These modules enhance multi-scale and frequency-aware feature

fusion, improving segmentation accuracy with fewer parameters and low computational cost [23].

In addition to the well-known U-shaped segmentation networks, several recent architectures have demonstrated notable advances in both accuracy and efficiency. DCSAU-Net introduces a deeper and more compact split-attention U-Net, effectively leveraging multi-scale feature fusion to improve performance on complex medical images [24]. H2-Former employs hierarchical hybrid transformers to capture both local and global dependencies efficiently, enhancing feature representation [25]. Swin-UMamba blends the Swin Transformer with Mamba-based modules and benefits from ImageNet pretraining, achieving higher accuracy with reduced computational cost [26]. U-KAN incorporates Kolmogorov-Arnold networks into the U-shaped architecture to strengthen nonlinear feature interactions [27]. Similarly, U-Recurrent Weighted Key-Value (RWKV) adapts the RWKV recurrent-transformer hybrid for segmentation tasks, showing the potential of sequence-model integration [28]. These approaches further illustrate the diversity of recent architectures that complement and extend U-shape-based frameworks.

III. METHODS AND MATERIALS

The neural network with symmetrical encoders and decoders is called 3D U-Net [29]. To provide higher-resolution features, the encoder and decoder in 3D U-Net combine features of same resolution using skip connections [30]. Furthermore, to accomplish feature extraction and restoration, the 3D U-Net structure design employs 3D-convolution, 3D-aximum pooling, and 3D-deconvolution as input for 3D medical image data [31]. With the use of this technique, segmentation accuracy can be increased by capturing the image's 3D spatial information. Attention mechanisms are used in several ways for medical image segmentation. To segment sublingual tiny veins, Yang *et al.* [32] suggested an interactive attention network that can instinctively recognize the target vein topology. The decision-level fusion network uses a modal image as only one input for singular segmentation-network, and the final segmentation result is obtained by combining the segmentation results of each network [33]. To get fusion features for training segmentation networks, input-level fusion-networks generally stack multimodal images in channel dimension [34]. Given that the network can preserve the original image information to the greatest degree while acquiring the inherent features of the image, this paper uses the input-level fusion-network to fully use the feature representation of multimodal images. In order to help the network focus more on important information, this work incorporates a dual channel cross-attention approach into the input stage fusion network. This method can fuse multimodal features and pay attention to the intricacies of the breast within them.

A. Improved Multi-graph Segmentation Algorithm

For Linux systems, related packages that are based on Python can be invoked directly for registration activities.

ANTs lower the registration constraints and 21 registration methods. The four values involved in the registration process are the transformation matrix from floating image to fixed image, the transformation matrix from fixed image to floating image, and the registration results from floating image registration to fixed image. To get the registration outcome from floating image registration to fixed image in the experiment, the quantity is utilised. Employ the default parameter combination and select the ‘SyN’ registration algorithm when utilizing the image analysis tools provided by ANTs. Real-time performance is critical to medical image segmentation. Deformable registration networks can thereby significantly reduce registration time while maintaining registration accuracy when deep learning techniques are used in place of conventional “precision” registration techniques. Fig. 1 illustrates the image registration architecture enhanced by depth learning. The fixed and floating images are delivered into the DL registration network in order to produce the registration deformation field. Through the operation of the spatial transformation network, the deformation field is applied to the tag image. Using the registration image and loss function, the network parameters are continuously altered to produce the full registration model. To prioritise RoI features, remove unnecessary features, and increase the target image block’s feature power, the Efficient Channel Attention (ECA) attention mechanism network is added to the basic registration network. As a result, the network extracts a feature map that is more noticeable; To expand the receptive field, introduce cavity convolution, and collect multi-scale data, 33×3 is an enhancement of the loss function. Navigate the image, change the network parameters, and improve the registration accuracy. Although the network model’s performance can be enhanced by adding attention modules, doing so will unavoidably make the model more complex. Wang *et al.* [35] proposed the ECA attention mechanism to balance complexity and performance. The primary changes made by ECA Net to SE Net were to propose a local channel mutual strategy that reduces dimensionality and can be efficiently implemented using one-dimensional convolution. The SE attention technique first uses channel compression to reduce the dimensionality and compress the input feature maps; however, this compression negatively affects the dependence relationships amongst the learning channels. As a result, the ECA attention mechanism employs one-dimensional convolution instead of dimensionality reduction to effectively accomplish extract dependence connections between channels and local channel interactions. In Fig. 1, its structure is displayed.

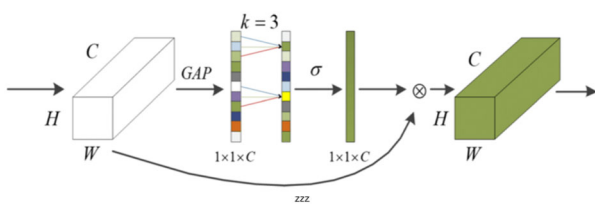


Fig. 1. ECA attention model.

The ECA attention mechanism has the least impact on the processing speed of the network model while guaranteeing the model’s performance and efficiency when compared to other attention methods. In order to improve the network and achieve performance optimisation more successfully and efficiently, the ECA attention mechanism is employed. This allows for the adaptive achievement of global correlation from both space and channel viewpoints in reaction to drop in image quality that occurs when U-Net networks down sample. This paper enhances the technique by using Dilated Convolution to address the problem of information loss [36]. Initially, some image segmentation issues were addressed by the proposal of cavity convolution. Nevertheless, the segmentation accuracy suffers, and image pixels are lost during the initial image reduction step. A hollow convolution was created to lessen this loss. The receptive field is commonly defined as the area on the feature map where a single pixel corresponds to its input map [37]. The receptive field expands as more feature and contextual information is collected. In contrast to the ordinary convolution, the hole convolution increases a hyperparametric hole rate (dilation rate), which is related to the number of inter cores, as shown in Fig. 2.

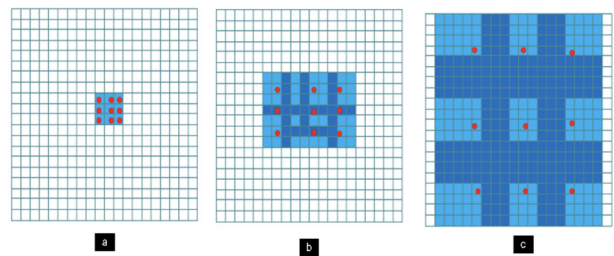


Fig. 2. Convolution of cavities and receptive fields 3×3 kernel with a) One void rate, b) Two void rate and became 7×7 kernel, c) Four void rate and became 15×15 kernel.

The benefit of dilated convolution lies in its ability to enable network model to acquire additional feature info for every convolution while preventing information-loss. Consequently, training accuracy and speed can be increased by employing hollow convolution to excerpt features from the image’s complete regions. Utilizing two channels, the experiment enhances the decoder’s variable performance by obtaining twice as many image features as the original U-Net. Fig. 3 depicts the network’s fundamental convolutional network topology, which is parameterized using a CNN that resembles U-Net. The network, which is a transcoder decoder with skip connections and comprises 4 down-sampling and 4 up-sampling structures, can create deformation fields from given fixed and floating images ϕ . Hole convolution is added at the bottom of the network to boost the receptive field and enhance the correlation between features. Reduced feature information loss during the down-sampling process is the goal of the introduction of cavity convolution. With an expansion rate of 2, the convolution nucleus will be filled with two zeros, the receptive field will be expanded, multiscale information will be obtained, features will be extracted from the image, and training speed and accuracy will be increased.

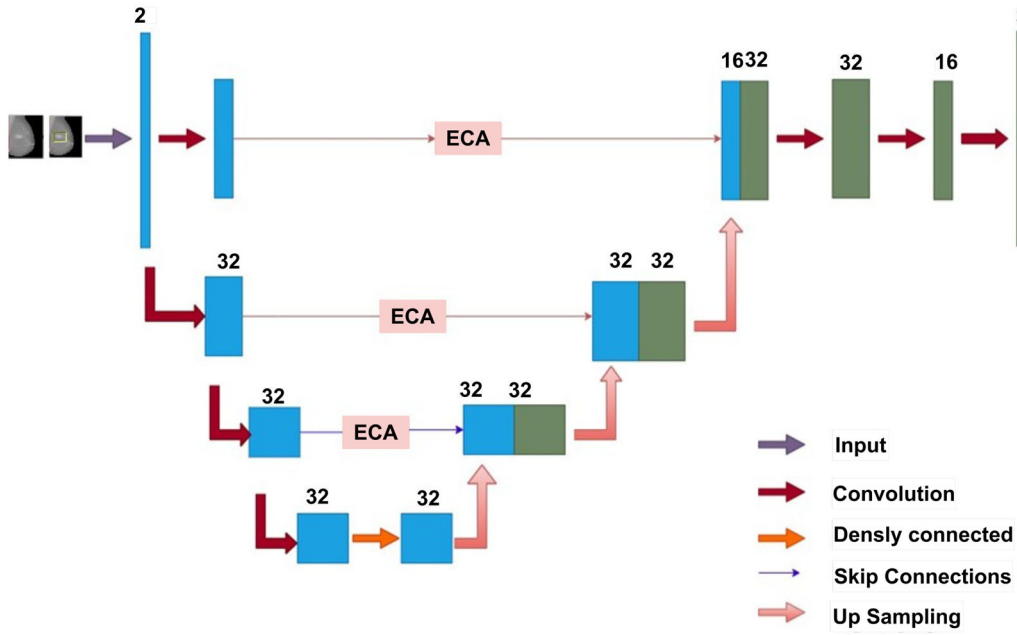


Fig. 3. Architecture of the proposed model.

The original coder contains four down-sampling modules. To increase the coder's ability to extract features, an ECA module has been added to every single one of the four down-sampling modules in this article after each down-sampling process. Once the original features were extracted and an ECA module was added, an ADD addition step was introduced before doing ECA processing. It was then linked to the feature layer that was up sampled to gain finer features and enhance the functionality of the model. EDD net may be used to model the function, input fixed and floating images, and forecast the displacement vector field end-to-end. During the training phase, the moving image M and the displacement vector field the training process find the maximum number of parameters by minimizing the phase loss among the curve image and the fixed image F as well as the regularisation loss of the displacement vector field. The curve image is obtained by the curve. As seen in Fig. 4, the Localization Network, Grid-Generator, and Sampler comprise the three components of a spatial transformer's operational mechanism.

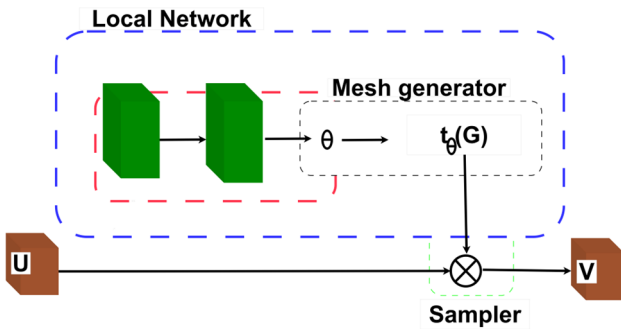


Fig. 4. Network for spatial transformation.

B. Principle of Rib Image Segmentation

The U-Net network, which consists of two stages, down

sampling and up sampling is variation of the end-to-end fully convolutional network. The down-sampling module uses two cascaded 3×3 and 2×2 to gradually extract deep spatial information. While up sampling simplifies deep feature maps, down sampling progressively extracts deep features. Furthermore, it is challenging to extract spatial information from shallow features. Pixel categorization and positioning are solved by deep features, and the output features from the down-sampling and up-sampling are combined and bonded as input for the subsequent module in up-sampling process. The application of shallow-level features is the key to segmentation for medical segmentation problems, where the form is comparatively stable without complicated semantics. The capacity to extract features at shallow levels is enhanced by jump structures. U-Net placement and segmentation outcomes are more accurate when symmetrical structures and skip connections are used.

The Squeeze and Excitation model (SE) is presented as a solution to the under-segmentation issue in finding tiny fracture contours created by complicated rib images, unfixed segmentation objectives, and lack of semantic richness. When low-level features lack semantic background, the SE module is implemented to gather high-level semantic information, figure out each feature channel's weights, amplify important features with weights, and stifle unimportant features to increase the accuracy of model segmentation. In Fig. 5, the input feature is represented by x , and the width, height, and number of channels are denoted by W , H , and C , respectively. Global average pooling converts each channel dimension into an average that represents the average information of the channel image. Next, by using two fully connected networks to excite each other, the weight values for each channel are produced (a number between 0 and 1 denotes the degree of relevance; the greater the number, the higher the proportion). Lastly, the

feature image x is multiplied by the channel weight value via the excitation function to process the scale weighting. By gaining knowledge of each channel in the feature image, SE structure distributes weight values. There are few calculations involved in the entire process. It will only marginally enhance the model's complexity and a few parameters. In the basic 3D U-Net, up to four subsamples can be used to obtain the receptive field. To expand the receptive field after the last subsample, the network incorporates a pyramid pooling module. The Pyramid Pooling Module (PPM) is integrated into AfroNet to enhance global context awareness by aggregating features from multiple spatial scales. It performs pooling at different grid sizes (e.g., 1×1 , 2×2 , 3×3 , 6×6), followed by convolution and upsampling, which are then concatenated with the original feature map. This enables the model to retain both global context and local detail—critical for segmenting lesions with varying sizes and shapes. The pyramid pooling technique, like the U-Net network's skip connection, produces multi-scale information. Four subsamples' convolution results are channel fused, then the input feature image is spliced in to produce the output results. To provide model with broader Receptive field in 3D space and enable multiple ranges as well as global feature information, the network employs pyramid pooling [38].

The 3D RSPU Net network, like 3D U-Net, makes use of identical symmetrical structures for encoding and decoding, with input photos processed through a preliminary $3 \times 3 \times 32$ feature extraction using 32 channels. To ensure that with a residual network, the segmentation effect is enhanced. the convolutional module instead as in Fig. 6. The section on encoding is made up of four module groups, each of which has an SE structure and a down-sampling module. The SE two convolution processes are included in the framework to increase the number of channels and greatly improve segmentation impact. The four module groups that make up the decoding part are up-sampling and an SE structure. The output of decoder's preceding layer and encoder's output data at appropriate location provides the input for up-sampling. Both the up and down sampling rates are two. Every time, there is a two-fold increase in the total number of down-sampling

channels and a two-fold loss in resolution. At the network resolution, a pyramid pooling module is inserted following three iterations of down-sampling. There are 4 parallel average pooling layers in the module. The Receptive field has dimensions of 1, 2, 3, and 6, in that order. To obtain more detailed background features, use a separate Receptive field. To achieve the splicing of encoder's shallow features and decoder's deep features, a skip connection is set apart between the two (Fig. 7).

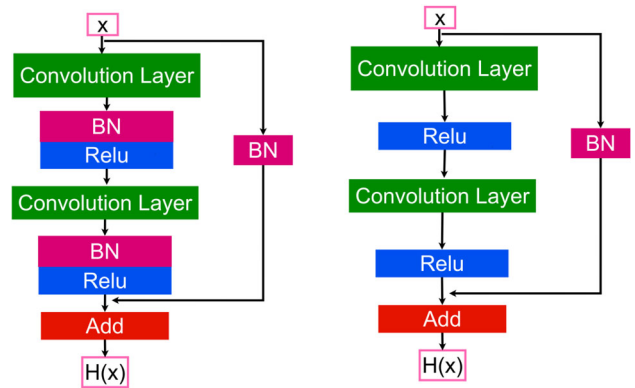


Fig. 5. Two types of residual modules: standard and enhanced.

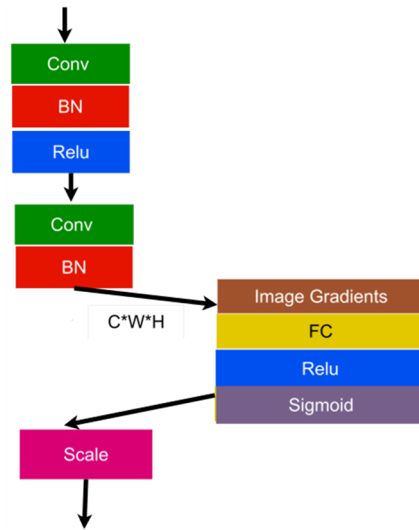


Fig. 6. Module of squeeze and excitement.

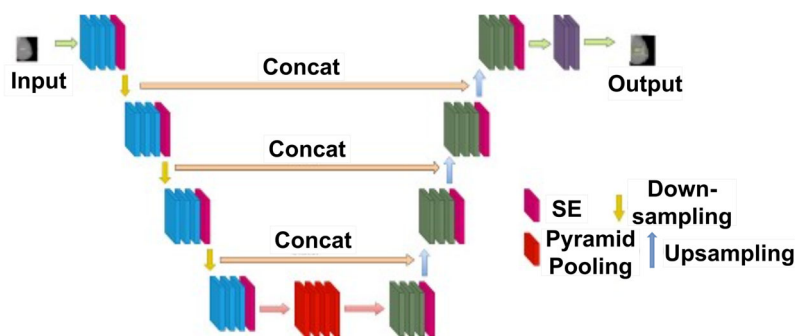


Fig. 7. Framework for 3D RSPU-Net.

C. Trans-UNet Network Architecture

An improved Trans-UNet network model that incorporates morphological structure learning is

developed using the Trans-UNet architecture, Transformer, and CNN structures as the benchmark networks on the network structure. Joint network learning

and morphological structure constraints form the basis of this three-dimensional medical image segmentation technique. The foundation for additional segmentation is laid by constructing a morphological structure constraint module using prior morphological structure information, extracting the shape information of the segmented object, and using a U-shaped Trans-UNet network structure to fully utilise the image’s high-level abstract features and low-level surface features. The encoder first uses CNN to extract features. Following each Res-Block module, representative attributes are further extracted from the CNN findings using maximum pooling and average pooling. The morphological structure module then receives these features. After the three-layer CNN feature extraction module, further features are extracted using a Transformer, which improves the features’ longevity and interpretability. This makes it possible to extract both local and global feature information. The decoder uses cascade-up sampling and skip connections to concatenate features from the same level. Fig. 8 depicts the network structure of the suggested approach. The benefits of this paradigm are primarily represented in two ways: first, the CNN and Transformer working together can help get the combined information on local and global features and raise the segmentation’s precision and accuracy; Creating a form for earlier information extraction comes in second module to limit the model’s final output, boost additional

educational data, increase accuracy, resilience and expansion of the segmentation model. A thorough explanation of the upcoming will be given below the two models’ organisational framework and method of application. Transformer uses a self-attention mechanism to assign weight to the image. It designates the background area as low weight and the target area as high weight. In this manner, CNN part of the network is employed for feature extraction, and the network concentrates more on learning to segment target area. The precision and accuracy of segmentation can be improved more when CNN and Transformer are combined. To produce superior segmentation results, a joint segmentation method is utilized in the segmentation task to segment numerous targets together. This method involves supplementing information between different tasks. By developing a module for extracting shape information from 3D medical images, the morphological structure of different organs, tissues, and structures may be retrieved. The edges of the organs, tissues, and structures that need to be segmented are recovered from medical images utilizing edge extraction techniques to obtain previous information. Convolution computations are used to produce the final shape information extraction results after they have been stitched together, which limits the U-Net model’s final output and increases the segmentation model’s accuracy, robustness, and generalization.

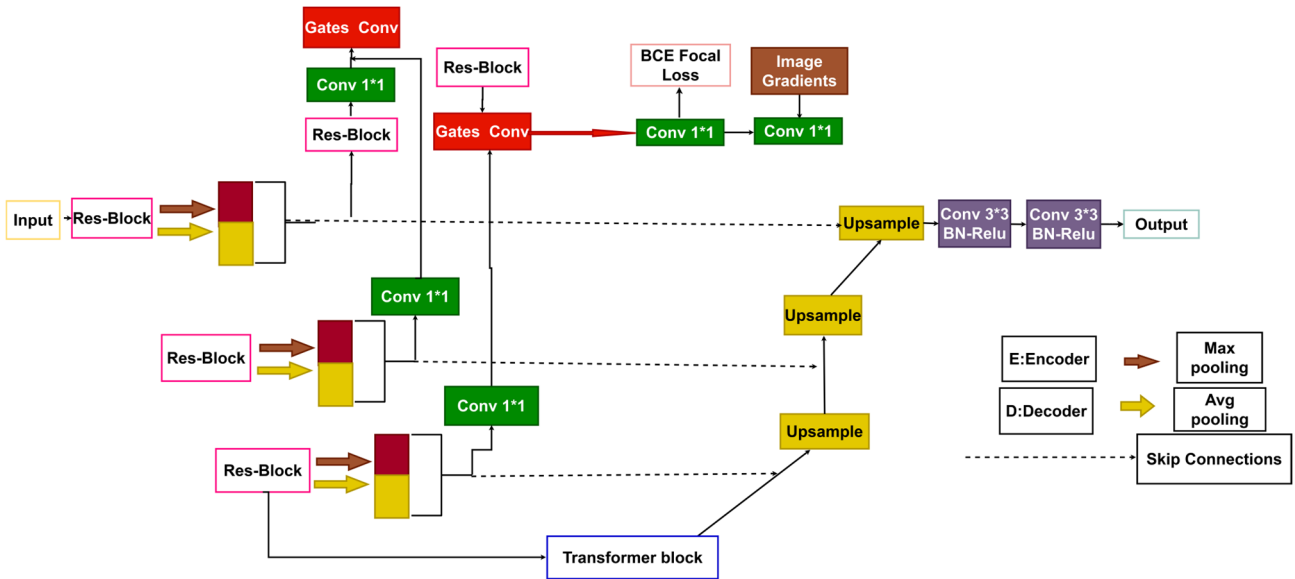


Fig. 8. Diagram of the proposed network structure.

IV. RESULTS

A. Dataset Description

More than 20,000 segmentation annotations of breast cancer tissue regions are present in the Breast Cancer Semantic Segmentation (BCSS) dataset, which is sourced from TCGA [39]. This dataset has been scaled to 224×224 and 512×512 pixels per image, respectively. While the 512×512-pixel version retains more detail and accuracy, the 224×224-pixel version seeks to promote the development of machine-learning models for tissue

segmentation with increased computing efficiency. Pathologists, residents, and medical students work together to annotate slides utilizing the Digital Slide Archive. This dataset provides a unique instructional opportunity to improve accuracy and experiment in data consumption. It is a fantastic resource for model training and slide analysis in study. In this research, we address new approaches to crowdsourcing for large-scale data production and emphasize the importance of this kind of data for semantic segmentation neural network training. Enhancing predictions for uncommon classes, combining it with other datasets, or finding new path omics and

genetic biomarkers are some creative uses for this dataset. Fig. 9 shows some sample copies of the BCSS dataset.

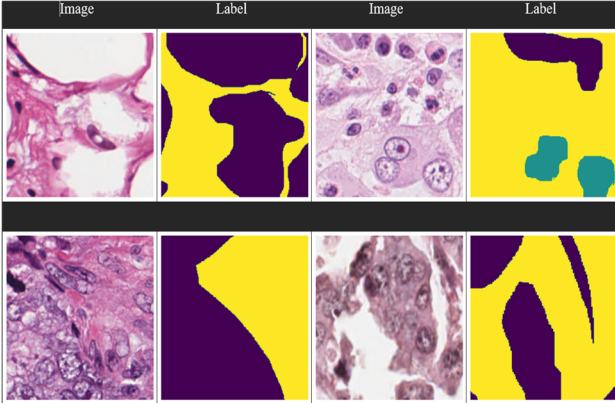


Fig. 9. Some samples of BCCS dataset.

B. Tools and Hardware

The studies in this paper were all conducted by means of open-source TensorFlow tool. Two NVIDIA GeForce GT1080ti graphics cards (8 GB of memory) are installed on Windows (64-bit) system. The network model's starting learning-rate is customary at 0.001, number of iterations for updating the weight is arranged at 50, and learning rate drops by half for every 10 changes to the weight. To enhance the assessment of the suggested network model, tests were carried out using one widely used medical breast image dataset: BCCS. The diabetes patients in the BCCS data set had different levels of damage to the white matter as in Fig. 9. This article's trial uses 10% of the data for validation, 10% for testing, and 80% of the data for training. Every image has gone through deviation correction, and medical professionals manually segment the target images.

The segmentation efficiency of the network for breast tissues is assessed using the three most widely used evaluation indicators to assess the efficacy and dependability of the algorithm presented in this research. The Sørensen-Dice Coefficient (DC) is one of the three indicators. Hausdorff Distance (HD) and Absolute Volume Difference (AVD). Their formulas are shown in Eqs. (1)–(3).

$$D_{dice} = 1 - \frac{2|P \cap G|}{|P| + |G|} \quad (1)$$

$$A_{AVD}(S, L) = \frac{VP - VG}{VG} \times 100\% \quad (2)$$

$$H_{HD}(P, G) = \max[h(P, G), h(G, p)] \quad (3)$$

The segmented image of the prediction model is represented by P , the segmented image of the real image is represented by G , volume of predicted segmentation results is represented by VP , and the volume of the segmented real image is represented by VG . Eqs. (4)

and (5) display the expressions for $h(P, G)$ and $h(G, P)$, respectively.

$$h(P, G) = \max\{p \in P\} \min\{g \in G\} |p - g| \quad (4)$$

$$h(G, P) = \min\{p \in P\} \max\{g \in G\} |g - p| \quad (5)$$

The accuracy of the segmentation increases with the size of the Sørensen-Dice Coefficient. To confirm the efficacy of the suggested modules, this paper runs tests using the BCCS dataset.

C. Discussions

Metrics like the Jaccard's Index and Dice Coefficient are castoff to measure efficiency of deep learning models especially when dealing with the semantic segmentation problem. Whereas the Dice Coefficient calculates the level of resemblance between mask and predicated, Jaccard's Index gauges the degree of overlap between mask and predicated. Fig. 10 shows the dice and Jaccard coefficients for individual classes of the proposed model.

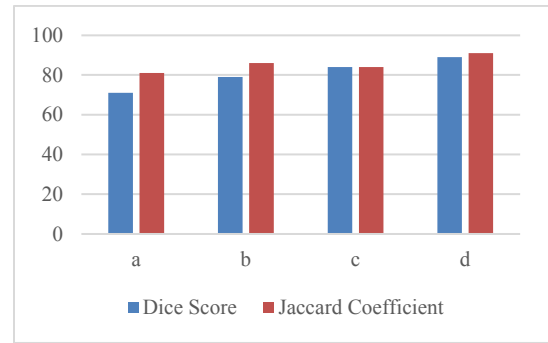


Fig. 10. Dice score and Jaccard coefficients obtained with a) 3D-U-Net; b) 3D-U-Net+MEM 3D; c) 3D-U-Net+MEM+DCRA; d) AfroNet.

From Fig. 10, the lowest Jaccard value occurred with U-Net and the highest dice occurred with the proposed model respectively. Similarity coefficients are frequently employed in the comparison of segmentation algorithms with a ground truth, or established reference mask. The ground truth, or actual tumor extent, is typically unknown when comparing two distinct imaging modalities to determine volume differences. As a result, it is important to interpret similarity coefficient values cautiously because they may be deceptive. Using 3D U-Net as the foundational network with no embedded MEM, DCRA, or IPD modules. The segmentation performance of 3D UNet has increased because of the gradual addition of MEM, DCRA, and IPD modules. Apart from the AVD indicator, all the 3D U-Net evaluation indicators have improved since the MEM module was added (which is portrayed as 3D U-Net + MEM in the network module). Of these, the Dice indicators improved by 4.89 and 6.13 percentage points, respectively. Six indicators are better than with 3D U-Net + MEM when the DCRA module (which is portrayed by the network module 3D U-Net + MEM + DCRA) was added to 3D U-Net + MEM framework. Of them, the Dice indications and Jaccard rose to 91.25% and 88.45%, respectively, from 72.28% and 81.25%,

respectively as shown in Fig. 10. This suggests that the attention module for dual channel cross-reconstruction developed in this work is capable of efficiently extracting features from several modalities, thereby enhancing the network's segmentation performance.

To verify the efficacy of the IPD module, a fourth set of tests was conducted using the 3D U-Net + MEM + DCRA + IPD framework. The 3D U-Net + MEM + DCRA + IPD framework produced the best segmentation results, as Table I demonstrates; as a result, the technique suggested in this article performs better in feature extraction and segmentation. Although the segmentation accuracy of the proposed segmentation network model is significantly higher than that of the 3D U-Net network, its rate of operation is lower because its parameter quantity is almost double that of the 3D U-Net network. The number of parameters needed for the model after including different modules and the performance of each 32×32 . The comparison of the runtime of 32 3D images is shown in Table II.

TABLE I. RESULTS OF ABLATION EXPERIMENTS ON THE DATASET

Model	AVD	HD
3D U-Net	6.25	3.12
3D U-Net + MEM	5.52	2.96
3D U-Net + MEM + DCRA	5.51	2.52
3D U-Net + MEM + DCRA + IPD	4.25	1.29
AfroNet (without PPM)	4.22	1.19
AfroNet (with PPM)	4.12	1.02

TABLE II. MODEL PARAMETERS AND RUNTIME

Models	Model Parameter Quantity/107	Running time/ms
3D U-Net	2.25	196
3D U-Net + MEM	3.35	258
Proposed model AfroNet	3.58	352

Table II shows that, although needing more parameters and taking longer to complete than 3D U-Net, AfroNet has the highest segmentation accuracy based on Table I data. Table I also shows the impact of PPM module on proposed AfroNet architecture. In the registration step, the enhanced network is compared in this study with resampling [40], ANTs [41], Voxel Morph [42], ViT-V-Net [43], and Trans Morph [44] methods. The network uses the pre-processed dataset as a training set to educate the network model that is being employed. The batch size is one, the regularisation parameter is 4.0, and the learning rate (lr) is $4e^{-4}$, and there are 1500 training sessions overall. To optimize, use the ADAM optimizer.

The breast dataset's partial registration findings are displayed in Fig. 11. The BCCS dataset registration results are shown in rows one and two, while the LPBA40 registration results are shown in rows three and four. The fixed image is displayed in Fig. 11(a), the floating image is displayed in Fig. 11(b), and the generated deformation field is displayed in Fig. 11(c), which also applies spatial transformation to the floating image and yields the final registration result. The distorted image acquired using the registration technique in this study is shown in Fig. 11(c), and the deformed image generated by applying the Voxel Morph approach for registration is shown in Fig. 11(d).

The red box in Fig. 11 for breast nerve tissue, middle breast tissue, and image edge breast tissue shows how the method used in this study can capture characteristics more precisely and produce a registration result that is closer to the fixed image.

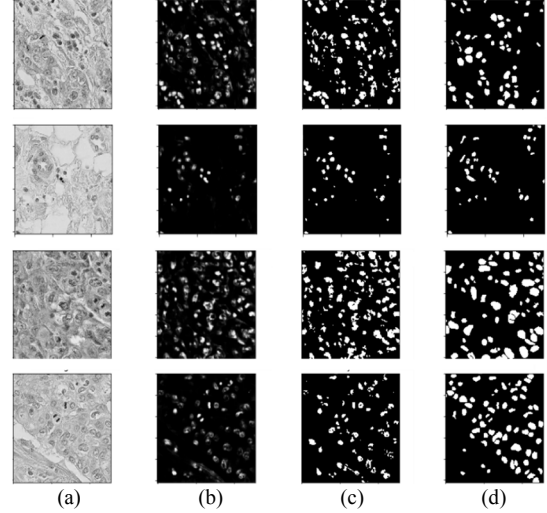


Fig. 11. Predictions with proposed model on BCCS dataset: From left to right; a) Original; b) Ground truth; c) 3D U-Net +MEM; d) AfroNet.

A registration network that can register images with notable internal structure differences is proposed in this study. While other methods can also perform spatial deformation on floating images, they sometimes suffer from insufficient local spatial transformation when dealing with local deformation issues. This makes it impossible to precisely register and deform floating images using the physical properties of fixed images, unlike the methods discussed in this article. This proves the efficacy of the algorithm proposed in this article. One important issue in evaluating deep and machine learning models is computational time. There is a direct proportionality between the time and the image patch size; that is, as the image patch size grows, so does the computational time. Accurate detection is increased by small patch sizes. Our suggested UMPNet architecture was trained on a parallel computing platform over nearly 12 h, and it required an average of 0.52 seconds for cell membrane and nucleus segmentation and classification on test data. The training loss (= 2.423) and validation loss (= 2.612) curves about iteration are displayed in Fig. 12(a).

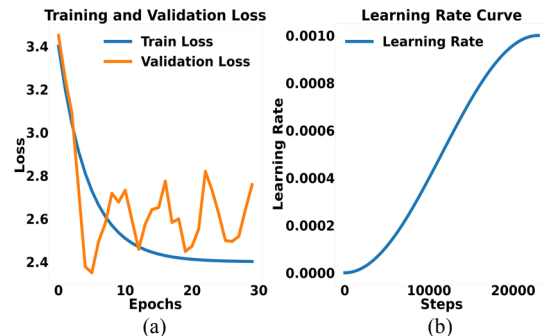


Fig. 12. Training dynamics of the model. a) Training and validation loss; b) Learning rate curve oversteps.

Selecting an appropriate learning rate is critical to the CNN models learning process. A low learning rate can cause the learning process to go slowly, whereas a high learning rate could cause the loss value to not converge, which would cause the learning process to fail. It is advised to employ “learning rate decay” to achieve the best results as shown in Fig. 13(b). This refers to lowering the learning rate as we loop through the training process, making the learning rate value a function of the current number of epochs. By doing this, we can obtain a learning algorithm that is faster without running the danger of it not convergent to a minimum loss value. The main statistic used to assess the accuracy of the image segmentation tasks’ outputs is mean Intersection over Union (mIoU). The model and the segmentation masks’ relationship to the ground masks will be evaluated. From Fig. 13(b), it is observed that, mIoU of the proposed model is drastically increasing with the number epochs. This shows our proposed model is fit well to our problem of semantic segmentation with the increased epochs.

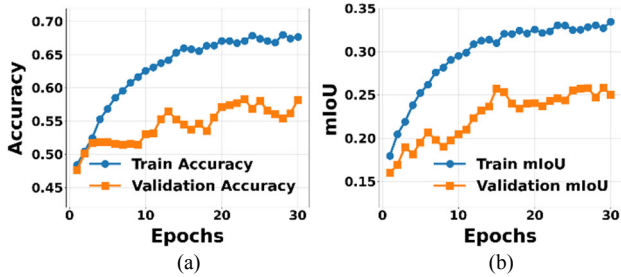


Fig. 13. Model performance metrics during training. a) Training and validation accuracy; b) Training and validation mIoU.

The aggregate frequency of correctly classified machine learning models is shown by accuracy. The accuracy of an ML model’s predictions for the target class is indicated by its precision. An ML model’s recall indicates whether it can find every object in the target class. Consider both the class balance and the costs of different errors while choosing the right metric. The recall, precision, and F1-Score of the suggested models throughout training with respect to epochs are displayed in Fig. 14.

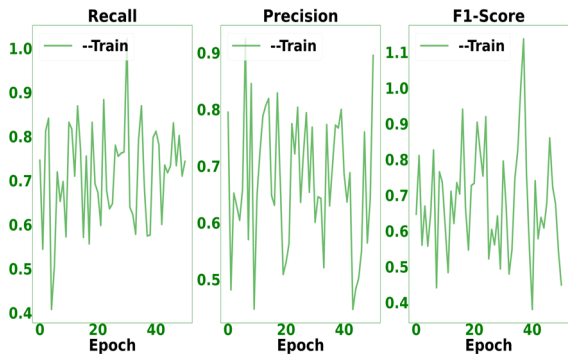


Fig. 14. Recall, Precision, and F1-Score of proposed models during training.

D. Loss Function

The difference between the forecasts and the actual is calculated using loss functions. Higher loss values suggest

that the model’s predictions are accurate, while lower loss values indicate that the model is more accurate. The loss function should always be minimised in the model; ideally, it should be close to zero. The trainable parameters, such as weights and biases, are learnt by models using the loss function and are modified in each iteration according to the gradient descent procedure. One loss function that is frequently employed for binary classification but occasionally ineffective for multi-class classification issues is cross entropy. It is derived from the mathematical notion of entropy, which is described as:

$$Entropy = P_i \log(P_i) \quad (6)$$

$$CrossEntropy = -\sum_i^n Y_i \log(P_i) \quad (7)$$

where P is the model-predicted output and Y is the ground truth or actual output.

We put a negative in front of the equation to make the entropy positive since the logarithm of probabilities, which may range from 0 to 1, yields a negative result. The weights will update in a way that increases the model’s confidence in forecasting the majority class while decreasing its concentration on the minority classes in the case of class imbalanced problems because the loss function followed by gradient descent primarily focusses on the majority class. Focus loss is the answer to this problem [44].

Focus loss: Focus loss ensures that forecasts on difficult examples improve over time rather than becoming overconfident with easy ones by focussing on the situations where the model fails rather than the ones where it can accurately predict. A technique known as “Down Weighting” in focus loss is used to achieve this. Down-weighting highlights the significance of challenging examples by reducing the influence of simple instances on the loss function. This approach can be implemented by adding a modulating factor (μ), $\mu = (1 - P_i)^\gamma$ to the Cross-Entropy loss.

$$Focalloss = -\sum_i^n \alpha (1 - P_i)^\gamma \log(P_i) \quad (8)$$

where cross-validation tuneable focusing parameter is named γ and α is weighing factor. The behaviour of Focal Loss for numerous values of γ and α is portrayed in Fig. 15. From Fig. 15, the following observations are made:

- 1) The loss function remains unchanged and turns into a cross-entropy loss since p_i is low for the sample that was misclassified and the μ is close to or exactly 1.
- 2) As the model’s confidence level increases, as shown by p_i equal to 1, the μ will tend to zero, lowering the loss amount for cases that are successfully classified. γ re-scales the μ so that the easy cases are down-weighted more than the hard ones, reducing the effect of the simple instances on the loss function. A γ value of 2 improves focus loss performance on our dataset.

- 3) When γ equals 0, focal loss and cross entropy are the same.

Recall loss: We employed a unique recall-based performance-balanced loss, known as recall loss, to address the imbalance dataset problem. The model weights are adjusted to minimise the loss function of that class based on the model’s recall value during training. It is an example of hard class mining rather than the hard example mining approach used in the focal loss. Unlike focus loss and other losses, the recall loss adjusts its weights dynamically during training according to the per-class recall value. The CB loss improves accuracy at the expense of Intersection-over-Union (IoU), which accounts for false positives in semantic segmentation. Our recall loss can effectively balance each class’s precision and recall, improve accuracy while preserve a competitive IoU.

$$Recallloss = -\sum_{i=1}^C \frac{FN_C}{FN_C + TP_C} N_C \log(P^C) \quad (9)$$

where N_C , FN_C , and TP_C represent multiple samples, class C false positive and true positive rates, respectively.

If there are other classes, the same formula applies. The recall-loss, which is weighted by the class-wise FN_C , can be used as the standard cross entropy. The second insight is that it is likely more difficult to classify minority classes, which have higher FN_C , than large classes, which have lower FN_C . Thus, like inverse frequency loss, gradients of majority classes will be suppressed while gradients of minority classes will be elevated [45].

Hybrid loss function: To improve model performance and training, we considered the benefits and drawbacks of the three loss functions listed above. As shown in Fig. 15, the loss function in our situation is the total of the three loss functions mentioned before.

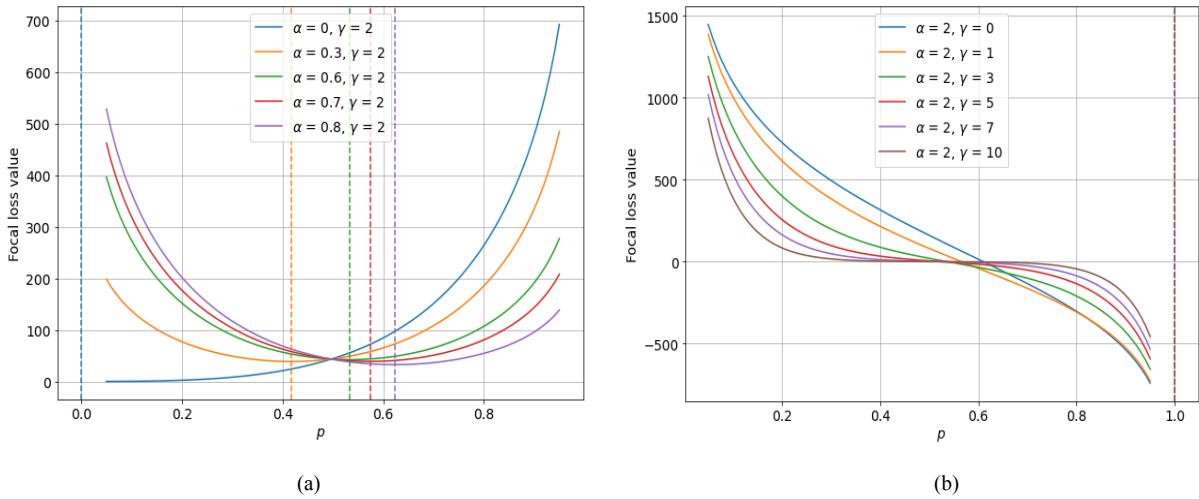


Fig. 15. Focal loss variation concerning a) α and b) γ .

E. Limitations

While AfroNet demonstrates strong segmentation performance, especially on high-quality and well-annotated breast cancer datasets, its effectiveness may diminish in cases involving poor image quality, significant staining variability, or extremely small tumor regions. The model is also sensitive to domain shifts across different imaging devices or institutions. Additionally, the current evaluation is limited to U-Net-based comparisons; broader benchmarking and real-time deployment in clinical settings remain part of our future work.

V. CONCLUSION

Since breast cancer is one of the major illnesses that many women experiences, it is vital to make the diagnosis process easier. The purpose of the suggested research model is to use ultrasound image processing to support clinical analysis of breast cancer. The system simulates a classifier model that receives as input the significant features derived from the AfroNet architecture’s layers. The suggested algorithm outperforms conventional diagnostic models by implementing suitable feature

selection and pre-processing techniques. The model’s accuracy is demonstrated by comparison with other comparable models, including 3D U-Net, 3D U-Net +MEM, 3D U-Net + MEM + DCRA, AfroNet techniques. The AfroNet algorithm demonstrated a maximum accuracy of 92.33%, surpassing all prior ensemble learning models customized for breast cancer diagnosis.

DATA AVAILABILITY

The data supporting the findings of this study can be obtained from the corresponding author upon reasonable request.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTION

Vuppula Manohar: Conceptualization, Methodology, Software, Writing—Original draft preparation, Visualization, Investigation, P. Srinivasa Rao: Supervision, Writing—Reviewing and Editing, Sreedhar

Kollem: Supervision, Visualization, and Investigation, Karri Chiranjeevi: Conceptualization, Methodology, B. Jaya: Visualization, Resources, M. Shashidhar: Methodology, Validation, Syed Musthak Ahamed: Supervision, Project administration, Appala Sruvan Kumar: Writing—review & editing, Manasa Koppula: Data Curation, Formal Analysis, Writing—Reviewing and Editing. All authors had approved the final version.

ACKNOWLEDGMENTS

The author wishes to express their gratitude to Vaagdevi Engineering College & Vaagdevi College of Engineering (Autonomous), CVR College of Engineering (Autonomous), SR University & Andhra University for their assistance.

REFERENCES

- [1] S. Sharmin, T. Ahammad, M. A. Talukder, and P. Ghose, "A hybrid dependable deep feature extraction and ensemble-based machine learning approach for breast cancer detection," *IEEE Access*, vol. 11, pp. 87694–87708, 2023. doi: 10.1109/ACCESS.2023.3304628
- [2] E. E. Fowler, C. M. Vachon, C. G. Scott, T. A. Sellers, and J. J. Heine, "Automated percentage of breast density measurements for full-field digital mammography applications," *Acad. Radiol.*, vol. 21, no. 8, pp. 958–970, 2014. doi: 10.1016/j.acra.2014.04.006
- [3] R. H. Lin, B. K. Kujabi, C. L. Chuang *et al.*, "Application of deep learning to construct breast cancer diagnosis model," *Appl. Sci.*, vol. 12, no. 4, p. 1957, 2022. doi: 10.3390/app12041957
- [4] B. A. Arnout, "The grief of loss among breast cancer patients during the COVID-19 pandemic: How can palliative care workers help?" *Work*, vol. 74, no. 4, pp. 1299–1308, 2023. doi: 10.3233/WOR-220400
- [5] S. Gupta and M. K. Gupta, "A comparative analysis of deep learning approaches for predicting breast cancer survivability," *Arch. Comput. Methods Eng.*, vol. 29, no. 5, pp. 2959–2975, 2022. doi: 10.1007/s11831-021-09679-3
- [6] R. Dandekar, A. Sharma, and J. Mishra, "A deep learning and feature optimization-based approach for early breast cancer detection," in *Proc. IEEE Int. Students' Conf. Electr., Electron. Comput. Sci. (SCEECS)*, 2024, pp. 1–7. doi: 10.1109/SCEECS61402.2024.10482104
- [7] L. Bouzar-Benlabiod, K. Harrar, L. Yamoun, M. Y. Khodja, and M. A. Akhloufi, "A novel breast cancer detection architecture based on a CNN-CBR system for mammogram classification," *Comput. Biol. Med.*, vol. 163, 107133, 2023. doi: 10.1016/j.compbiomed.2023.107133
- [8] S. Sengan, V. Priya, A. S. Musthafa, L. Ravi, S. Palani, and V. S. Swamy, "A fuzzy based high-resolution multi-view deep CNN for breast cancer diagnosis through SVM classifier on visual analysis," *J. Intell. Fuzzy Syst.*, vol. 39, no. 6, pp. 8573–8586, 2020. doi: 10.3233/JIFS-189174
- [9] S. Podda, R. Balia, S. Barra, S. Carta, G. Fenu, and L. Piano, "Fully automated deep learning pipeline for segmentation and classification of breast ultrasound images," *J. Comput. Sci.*, vol. 63, 101816, 2022. doi: 10.1016/j.jocs.2022.101816
- [10] L. D. U and M. T. R., "Analysis of deep learning and machine learning methods for breast cancer detection," in *Proc. Int. Conf. Comput. Sci. Emerg. Technol. (CSET)*, 2023, pp. 1–6. doi: 10.1109/CSET58993.2023.10346674
- [11] A. P. Windarto, A. Wanto, S. Solikhun, and R. Watrionthos, "A comprehensive bibliometric analysis of deep learning techniques for breast cancer segmentation: Trends and topic exploration (2019–2023)," *J. RESTI (Rekayasa Sist. Teknol. Inform.)*, vol. 7, no. 5, pp. 1155–1164, 2023. doi: 10.29207/resti.v7i5.5274
- [12] S. Pour, M. Esmaili, and M. Romozi, "Breast cancer diagnosis: A survey of pre-processing, segmentation, feature extraction and classification," *Int. J. Elect. Comput. Eng. (IJECE)*, vol. 12, no. 6, pp. 6397–6409, 2022. doi: 10.11591/ijece.v12i6.pp6397-6409
- [13] Aloyayri and A. Krzyzak, "Breast cancer classification from histopathological images using transfer learning and deep neural networks," in *Proc. Int. Conf. Comput. Recognit. Syst.*, 2020, pp. 491–502. doi: 10.1007/978-3-030-61401-0_45
- [14] T. Choudhary, V. Mishra, A. Goswami, and J. Sarangapani, "A transfer learning with structured filter pruning approach for improved breast cancer classification on point-of-care devices," *Comput. Biol. Med.*, vol. 134, 104432, 2021. doi: 10.1016/j.compbiomed.2021.104432
- [15] J. Mendes, J. Domingues, H. Aidos, N. Garcia, and N. Matela, "AI in breast cancer imaging: A survey of different applications," *J. Imaging*, vol. 8, no. 9, p. 228, 2022. doi: 10.3390/jimaging8090228
- [16] S. Shakya, K. Singh, and A. Saxena, "An adaptive deep learning technique for breast cancer diagnosis based on dataset," in *Proc. Int. Conf. Technol. Adv. Comput. Sci. (ICTACS)*, 2023, pp. 399–403. doi: 10.1109/ICTACS59847.2023.10390297
- [17] D. A. Zebari, H. Haron, D. M. Sulaiman, Y. Yusoff, and M. N. M. Othman, "CNN-based deep transfer learning approach for detecting breast cancer in mammogram images," in *Proc. IEEE Conf. Syst., Process Control (ICSPC)*, 2022, pp. 256–261. doi: 10.1109/ICSPC55597.2022.10001781
- [18] R. Poonia, V. K. Sharma, H. K. Singh, and S. Maheshwari, "Breast cancer detection: Challenges and future research developments," in *Proc. Int. Conf. Integr. Circuits Commun. Syst. (ICICACS)*, 2024, pp. 1–6. doi: 10.1109/ICICACS60521.2024.10498444
- [19] J. Vrdoljak *et al.*, "The role of AI in breast cancer lymph node classification: A comprehensive review," *Cancers*, vol. 15, no. 8, p. 2400, 2023. doi: 10.3390/cancers15082400
- [20] J. Mendes and N. Matela, "Breast cancer risk assessment: A review on mammography-based approaches," *J. Imaging*, vol. 7, no. 6, p. 98, 2021. doi: 10.3390/jimaging7060098
- [21] M. A. Choukali, M. C. Amirani, M. Valizadeh, A. Abbasi, and M. Komeili, "Pseudo-class part prototype networks for interpretable breast cancer classification," *Sci. Rep.*, vol. 14, no. 1, 10341, 2024. doi: 10.1038/s41598-024-60743-x
- [22] L. Sacca *et al.*, "Promoting artificial intelligence for global breast cancer risk prediction and screening in adult women: A scoping review," *J. Clin. Med.*, vol. 13, no. 9, p. 2525, 2024. doi: 10.3390/jcm13092525
- [23] L. Li, Q. Liu, X. Shi, Y. Wei, H. Li, and H. Xiao, "UCFilTransNet: Cross-filtering transformer-based network for CT image segmentation," *Expert Syst. Appl.*, vol. 238, 121717, 2024.
- [24] Q. Xu, Z. Ma, N. He, and W. Duan, "DCSAU-Net: A deeper and more compact split-attention U-Net for medical image segmentation," *Comput. Biol. Med.*, vol. 154, 106626, 2023.
- [25] A. He, K. Wang, T. Li, C. Du, S. Xia, and H. Fu, "H2Former: An efficient hierarchical hybrid transformer for medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 42, no. 9, pp. 2763–2775, 2023.
- [26] J. Liu *et al.*, "Swin-UMamba: Mamba-based U-Net with ImageNet-based pretraining," arXiv Preprint, arXiv:2402.03302, 2024. doi: 10.48550/arXiv.2402.03302
- [27] C. Li *et al.*, "U-KAN makes strong backbone for medical image segmentation and generation," in *Proc. of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 5, 2024, pp. 4652–4660.
- [28] S. Wang *et al.*, "U-RWKV: RWKV-based U-shaped foundation model for medical image segmentation," arXiv Preprint, arXiv:2510.07041, 2025. doi: 10.48550/arXiv.2510.07041
- [29] S. Zheng *et al.*, "MDCC-Net: Multiscale double-channel convolution U-Net framework for colorectal tumor segmentation," *Comput. Biol. Med.*, vol. 130, 104183, 2021.
- [30] X. X. Yin, S. Hadjiloucas, Y. Zhang, M. Y. Su, Y. Miao, and D. Abbott, "Pattern identification of biomedical images with time series: Contrasting THz pulse imaging with DCE-MRIs," *Artif. Intell. Med.*, vol. 67, pp. 1–23, 2016.
- [31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778.
- [32] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 2980–2988.
- [33] X. X. Yin, L. Yin, and S. Hadjiloucas, "Pattern classification approaches for breast cancer identification via MRI: State-of-the-art and vision for the future," *Appl. Sci.*, vol. 10, no. 20, p. 7201, 2020.
- [34] N. Tawfik, H. A. Elnemr, M. Fakhr, M. I. Dessouky, and F. E. Abd El-Samie, "Hybrid pixel-feature fusion system for multimodal

- medical images,” *J. Ambient Intell. Humaniz. Comput.*, vol. 12, no. 6, pp. 6001–6018, 2021.
- [35] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, “ECA-Net: Efficient channel attention for deep convolutional neural networks,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 11531–11539.
- [36] T. Peng, C. Wang, Y. Zhang, and J. Wang, “H-SegNet: Hybrid segmentation network for lung segmentation in chest radiographs using mask region-based convolutional neural network and adaptive closed polyline searching method,” *Phys. Med. Biol.*, vol. 67, no. 7, 075006, 2022.
- [37] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, “nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation,” *Nat. Methods*, vol. 18, no. 2, pp. 203–211, 2021.
- [38] X. Liu, L. Song, S. Liu, and Y. Zhang, “A review of deep-learning-based medical image segmentation methods,” *Sustainability*, vol. 13, no. 3, p. 1224, 2021.
- [39] M. Amgad *et al.*, “Structured crowdsourcing enables convolutional segmentation of histology images,” *Bioinformatics*, vol. 35, no. 18, pp. 3461–3467, 2019.
- [40] Z. Xu and L. He, “A medical image segmentation model integrating multiscale features,” *Comput. Knowl. Technol.*, vol. 19, no. 7, pp. 35–37, 2023.
- [41] Y. Ma, H. Hao, J. Xie *et al.*, “ROSE: A retinal OCT angiography vessel segmentation dataset and new model,” *IEEE Trans. Med. Imag.*, vol. 40, no. 3, pp. 928–939, 2020.
- [42] J. Liu, X. Ni, and Y. Li, “Research on medical cell image enhancement algorithms based on image segmentation,” *Internet Things Technol.*, vol. 13, no. 2, pp. 40–41, 46, 2023.
- [43] W. Li and C. Wu, “Medical image segmentation network based on multi-level residuals and multi-scale,” *J. Hubei Univ. Technol.*, vol. 38, no. 1, pp. 38–42, 2023.
- [44] P. Xu, “Research on medical image segmentation method based on U-Net and clustering,” M. S. thesis, Nanjing Univ. Posts Telecommun., Nanjing, China, 2021.
- [45] Y. Patel, G. Tolia, and J. Matas, “Recall@k surrogate loss with large batches and similarity mixup,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2022, pp. 7502–7511.

Copyright © 2026 by the authors. This is an open access article distributed under the Creative Commons Attribution License ([CC-BY-4.0](https://creativecommons.org/licenses/by/4.0/)), which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.