International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 17 (2025) pp. 26–41

Article

A Deep Learning Model for Accurate Multi-Modal Feature Fusion and Segmentation of Breast Cancer Nodules

Kudithikunta Anusha [1,*] and Konduru Reddy Madhavi [2]

[1] Research Scholar, School of Computing, Mohan Babu University, Tirupati 517102, Andhra Pradesh, India
[2] Professor, AI&ML, School of Computing, Mohan Babu University, Tirupati 517102, Andhra Pradesh, India; kreddymadhavi@gmail.com
* Correspondence author: anushagsk@gmail.com

Received date: 18 July 2024; Accepted date: 21 September 2024; Published online: 6 January 2025

Abstract: Breast cancer is characterized by the uncontrolled growth of breast cells, resulting in the formation of tumors. In this study, we propose the AMFFR-Net model, a deep learning network capable of performing multi-modal fusion and segmentation of breast cancer nodules with medical images in specific MRI, ultrasound, and mammograms. The proposed AMFFR-Net is an advanced deep neural network that can enhance the diagnosis process by utilizing multi-modal feature fusion and refinement methods to achieve precise segmentation of the regions of interest. Quantitative assessments using diverse datasets show that AMFFR-Net is superior to other baseline models in terms of different criteria. On MRI datasets, it shows an outstanding AUC score of 98.94%, an F1-score of 94.8%, and an overall accuracy of 97.0%, demonstrating that it is superior in terms of discrimination and provides more accurate delineation of anatomical structures. Such high performance is also replicated in the ultrasound and mammogram datasets, where AMFFR-Net predicts higher AUC, F1-score, and Dice coefficient values with lower HD95 values indicating accurate boundary detection.

Keywords: multi-modal feature fusion; segmentation; mammograms; AMFFR-Net

1. Introduction

Globally, the number of cases of breast cancer (BC) among women has experienced a substantial rise compared to the past few years. It has been the most frequently diagnosed cancer among women and a leading cause of mortality due to cancer, which comes in second when it comes to all causes of mortality for females. The International Agency for Research on Cancer (IARC), in its 2020 world cancer report, reported that cancer is the leading or the second leading cause of death for 134 out of 183 countries for people within the age group of 30–69 years [1]. Lung cancer remains the leading cause of cancer deaths for both men and women; breast cancer is most prevalent in women, and prostate cancer is most common in men. The IARC predicts that the number of cancers will more than double from 19. In a single year, from 2017 to 2018, the number of tourists grew from roughly 1 million to 29. According to the findings, there will be up to 5 million deaths by 2040, and fatalities may double under the assumed conditions. 6 million to 16.4 million [1].

Among all types of cancer, breast cancer occupies the second-highest rank in terms of prevalence and mortality for women. It presents a vast array of medical and economic problems. This disorder has a variety of dimensions in terms of its impact on women's health, encompassing both the physical, mental, and social fields of wellbeing. The fact that it acts rapidly, restlessly, and early in finding out and diagnosis is very significant, not only for the better survival rates but also for reducing the costs associated with the treatment. Identifying early-stage cancer can increase the survival rate by 85 percent. One of the cells in the breast structures, most commonly in the ducts or the milk-producing lobules, forms

breast cancer. It can later infiltrate nearby tissues and extend to other parts of the body. The stage of breast cancer shows how much the malignant cells have already spread from the area of the original tumor [2]. The genetic flaw that develops as a result of aging or general body breakdown is the primary cause of 85%–90% of cancer cases. Although it's sometimes hard to identify the exact factor that causes the damage, breast cancer originates within cells as a result of DNA damage or mutations that happen to the cells. Three primary types of breast cancer are invasive carcinoma, ductal carcinoma, and invasive lobular carcinoma. The majority of breast carcinoma cases are invasive carcinoma, which originates from the breast ducting system and subsequently spreads into surrounding tissues, and invasive lobular carcinoma, which develops from the breast glandular tissue [3]. What those advancements in imaging technologies, imaging modalities, and computational development have done has put a new emphasis on breast cancer research. Medical imaging has now become an indispensable tool in breast cancer care, serving as one of the main diagnostic methods from screening to treatment decisions and post-treatment monitoring. On the other hand, progress has also played a cardinal role in the invention and incorporation of computer-aided detection systems. When the patient is initially consulted with the symptoms affecting the breast, the physician most likely initiates a diagnostic evaluation that involves the use of mammography (an X-ray specializing in breast imaging). Often, women with dense breast tissue or those with a predisposing factor, such as a family history of breast cancer, receive an inconclusive diagnosis. This matter would need further diagnostic procedures to arrive at an accurate diagnosis. Radiologists usually use breast ultrasounds (ultrasonography) to assess lumps that are not easily visible through mammography in women with dense breast tissue [4].

1.1. CAD (Computer Aided Design)-Based Imaging Modalities to Predict Brest Cancer

Obtaining a precise and correct diagnosis of breast cancer (BC), among other things, reduces the disease's risk and improves survival rates. This simply means that early detection of breast cancer leads to early treatment initiation. Regularly emerging imaging modalities enable early diagnosis of the disease's onset. Realizing the essence of this matter, medical imaging is a valuable means of cancer recognition and diagnosis. Healthcare professionals use these modalities for various purposes, including screening, diagnostic evaluation, and ancillary testing to boost their confidence and improve diagnostic accuracy. Moreover, Figure 1 summarizes the imaging techniques used for BC diagnosis today, which include digital mammography, ultrasonography, histopathology, magnetic resonance imaging, computed tomography, PET-CT, and thermography [5]. Each of these separately mentioned markers holds particular importance and provdies a specific profile for the diagnostic process.

- Mammography
- Ultrasound
- MRI

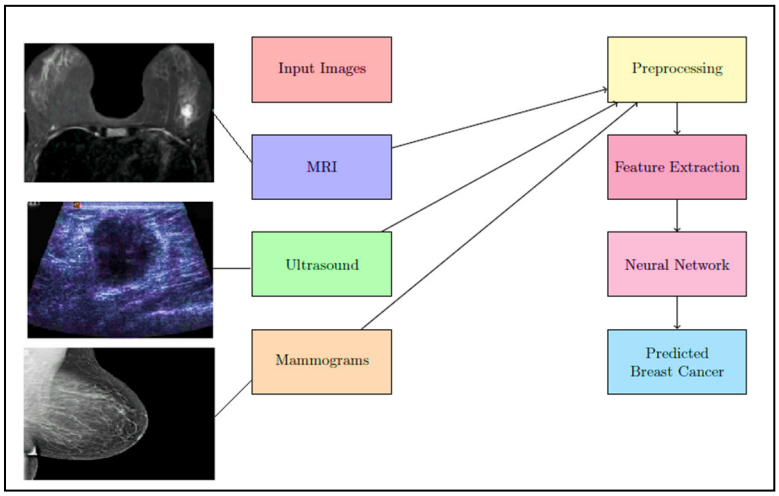

Figure 1. CAD-based Imaging Modalities to Predict Brest Cancer.

1.2. Role of Deep Learning in Breast Cancer Diagnosis

The role of DNNs in breast cancer detection cannot be overstated; the AI-equipped machine can be leveraged to examine imaging datasets. DNNs possess superior imaging analysis skills for diverse

imaging modalities such as mammograms, MRIs, and ultrasounds, and they help to disclose the inscrutable characteristics of breast malignant entities. Artificial intelligence's analytical power empowers DNNs to pattern out the best-hidden traits and irregularities within images that are unseen by the naked eye. Consequently, DNNs enable precise diagnosis and timely treatment when the likelihood of effective treatment is at its peak. Aside from DNNs' potential to automate the screening process with their ability to swiftly analyze huge volumes of medical images, they can also conduct deep analysis of implanted medical devices, such as implantable pacemakers. Through the use of such automated technology, healthcare professionals may be able to free up some of their time and resources and enhance the overall productivity of mammography screening. Another important role of the DNNs is to contribute to precision medicine by integrating imaging, genomic, and patient-specific data, such as past medical and pharmacological influences [6]. For example, one of the most essential conditions for AI advancement is access to different, powerful, and labeled datasets for training the models. The question of privacy, although it may have a significant effect on DMN usability, could become a major obstacle when it comes to medical imaging data security. Furthermore, the widespread reference to DNNs as "black-box" models poses a challenge to human understanding and explanation. The clarity of DNNs' analogical thinking is critical in order to gain the trust and acquiescence of medical practitioners and patients. Training models on a single dataset may limit the effectiveness of DNNs when applied to different populations and/or imaging modalities, which is another critical disadvantage to their generalizability.

Deep neural network (DNN) models must possess robustness and adaptability to gain widespread acceptance across various settings and clinical contexts. In specific, the main challenges these methods possess hinder the application of DNNs in clinical practice, despite their significant merit in replacing traditional low-specific diagnostic methods for early breast cancer detection. Therefore, we must carefully address these challenges, prioritizing safety, effectiveness, and ethical considerations. Researchers' efforts in collaboration with healthcare providers, policymakers, and stakeholders are necessary to shape DNNs into a tool that can be transformative for breast cancer treatment and patient care development [6].

1.3. Motivation

Breast cancer is still one of the most widely spread and deadly types of cancer among women around the world, obscuring all the other problems. The traditional methods of diagnosing breast cancer, like mammography and biopsy, have had proven results, but in general, they are not devoid of any deficiencies. For example, mammography screening may give a positive result that the woman does not have cancer, whereas she really has a very mild cancer, which causes anxiety and further invasive investigation. To cope with these challenges, it is observed that there is an increase in the number of research studied considering the use of artificial intelligence and deep learning to improve breast cancer detection. Recent research studies experiments the usage of AI and DL algorithms and methods that take breast imaging data features into account, attempt to make the performance of diagnostic systems more sensitive, specific, and overall accurate.

The outlined approaches in this article address major challenges related to conventional breast cancer detection modalities. The proposed method, which uses the multiscale self-attention mechanism, adaptive feature fusion, and refinement techniques, will greatly increase the accuracy of cancer lesions identification and localization from a medical image. Second, the proposed method tackles the harmful effects of highly variable imaging data and unreliable diagnostic outcomes, equipping healthcare service providers with novel tools for faster disease identification.

The main contributions of the article include:

- Development of Multi-Scale Self-Attention Module (MSSAM) convolutional layers, including multi-scale with different kernel sizes, Multi-scale Channel Attention Module (MSCAM), and Adaptive Spatial Attention Module (ASAM). This method offers an innovative feature extraction and attention-guided segmentation of various breast cancer modalities.
- Designing an Adaptive Feature Fusion and Refinement (AFFR) module to achieve the goal of feature fusion, transformation, and refinement, which provides a good effect on the uniting of many features and improves the segmentation accuracy. The feature deformation and confidence-based feature refinement methods prolong the feature extraction process, and, hence, classifying the unsolved segmentation is much more precise.
- Integrates Binary Cross Entropy (BCE) loss function with proposed algorithm to perform the comparison of predicted and ground truth labels directly.

2. Related Works

There are different imaging methods, like digital breast tomosynthesis (DBT) and breast ultrasound (US), that can be used together with AI algorithms to find breast cancer. This is called multimodal fusion [7]. This technique aims to enhance precision by utilizing ML algorithms trained on diverse data types, such as the CBCD [7]. The system, utilizing UDF and ML algorithms, can achieve enhanced sensitivity and specificity when identifying breast lesions, particularly Researchers have found that this fusion method can greatly lower the number of false alarms and improve the accuracy of identifying cancerous lesions. As a result, breast cancer can be found early and correctly [7]. Furthermore, the combination of a logistic regression method with statistical processes yielded remarkable results, demonstrating the highest degree of accuracy in early breast cancer detection [8].

The multispectral fusion enhances breast cancer detection accuracy by combining various data sources. We employ various techniques, including DNA methylation, copy number alterations, and 4-nucleotide oligomer end motifs, in building a model as mentioned in Table 1. This results in a multi-feature machine learning model that is able to achieve high accuracy in the early diagnosis of breast cancer [9]. Multimodal analysis, in contrast, uses data such as genetic signatures, imaging, and clinical features to develop a feature-stacked classification model that increases accuracy and makes fast decision-making for doctors possible. The fusion of local binary pattern features using the Binary Firefly algorithm from ultrasound elastogram and echogram images has a remarkable effect on diagnostic performance, surpassing some known detection techniques [10]. Combining features from mammograms and digital breast tomosynthesis views, made possible by SVM and KNN systems, also makes it easier to find breast cancer.

These tools primarily generate deep multi-modal fusion networks (DMMFN), Integrated Stacking Classifiers, ultrasound fusion imaging systems, and multimodal affinity fusion networks (MAFN). These methods carry out feature extraction, weight coefficient adjustment, and fusing data from diverse modalities to enhance patient care, treatment decisions, and diagnostic accuracy in breast cancer management.

Table 1. Overview of related works.

Author	Problem Statement	Proposed Work	Results	Limitations
Gohariyan et al. [11]	Enhance breast cancer diagnosis using MG and MRI fusion	Pixel-level fusion with Gabor filters, Recurrent Fuzzy Neural Network, and SVM	Accuracy: 98.14%, Sensitivity: 95%, Specificity: 100%	Evaluated on a private dataset; potential variability with larger or public datasets
Hopp et al. [12].	Improve CAD system by combining MRI with MG (CC and MLO)	Image registration, feature extraction using GLCM, GLRM, and morphology, RF	Sensitivity increased by 2% (MRI+MG: 77.9% vs. MRI alone: 76.1%)	Limited sensitivity improvement; private dataset used for evaluation
Salem et al. [13]	Combine MG and MRI for better diagnostic performance	Visual primitives and texture patterns, score-level fusion with similarity measures	Accuracy: 80%, Precision: 66.66%, Specificity: 66.66%	Small private dataset; limited number of cases; low precision and specificity
Yuan et al. [14].	Distinguish malignant and benign lesions using MG and MRI	Gradient and morphological features for MG, texture for MRI, serial feature fusion, BANN	AUC: MG: 74%, MRI: 78%, Fusion: 87%	Private dataset; complexity in feature selection and fusion process
Lesniak et al. [15]	Discriminate between masses and normal tissue using MG and MRI	Extract texture, morphology, kinetics, gradient features, linear and non-linear SVM	AUC: MG: 76%, MRI: 73%, Fusion: 81%	Private dataset; moderate improvement in AUC for fused modalities
Ameur et al. [16]	Detect breast cancer using contrast-enhanced MG and MRI	Extract texture and shape features, serial feature fusion, ANN	AUC: MG: 96%, MRI: 97%, Fusion: 96%	Small private dataset; potential overfitting due to high accuracy on a limited sample

Bhooshan et al. [17]	Enhance lesion classification using MG, MRI, and Ult fusion	Extract multiple features (shape, edge, margin, texture, kinetics), LDA, BNN	AUC: MG: 76%, Ult: 81%, MRI: 73%, MG+Ult: 87%, Ult+MRI: 85%, MG+MRI: 89%, Triple Fusion: 95%	Private datasets; complex fusion process; need for integration of three different modalities

Significant research efforts have involved the creation of multimodal fusion systems for the improvement of the diagnosis of breast cancer.

Gohariyan et al. [11] have put forward a model that combines MG and MRI data with the mode of processing at the pixel level. In this system, Gabor filters are used for feature extraction, and a recurrent fuzzy neural network (RFNN) is also used in the process, along with SVM for classification. In the data set of 216 cases evaluated, the system demonstrated a pretty high accuracy rate of 98.95% and 95%, while the sensitivity and specificity rates were 14% and 100%, respectively.

Hopp et al. [12] presented the CAD system, which was comprised of two-view MG X-rays registered with MRI images of the intervertebral discs (CC and MLO views). Texture imaging was performed using the Gray Level Co-occurrence Matrix (GLCM) for MRI and a blended technique of GLCM, Run-Length Matrix (GLCM), and morphological features for myasthenia gravis. The feature-level fusion technique fused the extracted features, and a Random Forest classifier performed the classifier. It turned out that our proposed protocol brought approximately a 2% enhancement in sensitivity when we used both types of MG views along with MRI, with an average sensitivity of about 77%. undefined undefined.

Salem et al. [13] centered on the combination of MG with MRI based on three different visual parameters: size, shape, and segmented shape edge, which they modified by texture patterns. They employed a hybrid method at the score level that was based on similarity as a feature and achieved 80% accuracy and 66% precision. Prevalence of 66% and specificity of 66%. Accuracy is 66% on a private dataset with 35 mammogram images and 134 MRI volumes.

In an ongoing research study, Yuan et al. [14] implemented a CAD system which uses combined breast MRI and MG features to discriminate malignant from benign lesions. They used gradient and morphometric features for MRI and size, circularity and irregularity features describing lesions. The serial fusion was then performed and followed by the elimination of non-significant features (step-wise linear feature selection) and the classification of the data with the help of a Bayesian Artificial Neural Network (BANN). The best outcome (AUC was equal to 87%) was shown for a multi-modality fusion, compared to MG (74%) and MRI (78%) separately.

Lesniak et al. [15] proposed a CAD method that gives numerous features including the degree, morphologic, dynamic, and gradient properties from the MG and MRI. The linear SVM was used for feature selection, whereas, as the second step, a non-linear SVM was used for classification. The method achieved AUCs of 76%, 73%, and 81% for MG, MRI, and their combining models , respectively.

Amour et al. [16] investigated a framework for MG and MRI feature extraction from the textures and shapes presenting in contrast images. The serial method was used to implement the feature level at the time when it was possible for instant period classification using the Artificial Neural Network. Such a strategy yielded AUCs of 97%, 96%, 96% for MRI, MG and their fusion, respectively.

Bhooshan et al. [17] studied a coherent merging approach which comprised combined neuroimaging modalities of MG, MRI, and UL. They broke-down shape, edge, margin, and image texture from MG and Ult, and intensity, texture, morphology, and kinetic features from MRI. Bearing in mind that Linear Discriminant Analysis (LDA) was used within our system for feature selection and a Bayesian Neural Network (BNN) was used for classification, the system had AUCs equal to 76%(MG), 81%(Ult), 73%(MRI), reaching at least 95% for triple modality fusion and yields higher numbers for dual.

The discussed studies underscore the enhanced performance of breast cancer detection systems when fusing MG with MRI compared to using each modality independently. The results suggest that feature-level fusion using the serial method often yields better diagnostic performance. However, the use of small or private datasets and the integration of images from different patients highlight the need for more standardized and larger datasets to improve the generalizability and clinical acceptance of these systems. Future research should focus on robust feature extraction techniques and methods that can effectively capture the intrinsic relationships between different imaging modalities.

3. Proposed Mechanism

The AMFFR-Net architecture incorporates three key modules: the Multi-Scale Self-Attention Module (MSSAM), the Adaptive Feature Fusion and Refinement (AFFR) module, and the Adaptive Spatial Attention Module (ASAM). The MSSAM leverages multi-scale features and channel attention

mechanisms to guide the network in selecting informative features, capturing intricate details within various modalities of breast cancer and improving segmentation accuracy. The AFFR module integrates feature fusion, feature transformation, and feature refinement sub-modules. It aligns multi-modal features, aggregates them at the pixel level, and refines uncertain regions to produce precise segmentation results. These sub-modules ensure comprehensive and accurate representation of cancer nodules, even in the presence of spatial misalignments and uncertainties.

3.1. Multi-Scale Self-Attention Module

MSSAM comprises three major parts. i.e., convolutional layer with multiple sizes kernel, Multi-scale channel attention module (MSCAM) and adaptive spatial attention module (ASAM). Moreover input MSSAM feature maps processed through 3 parallel Conv (Convolutional layers) in order to obtain the 3 feature maps with multiple receptive fields. Figure 2 shows the MSSAM module architecture.

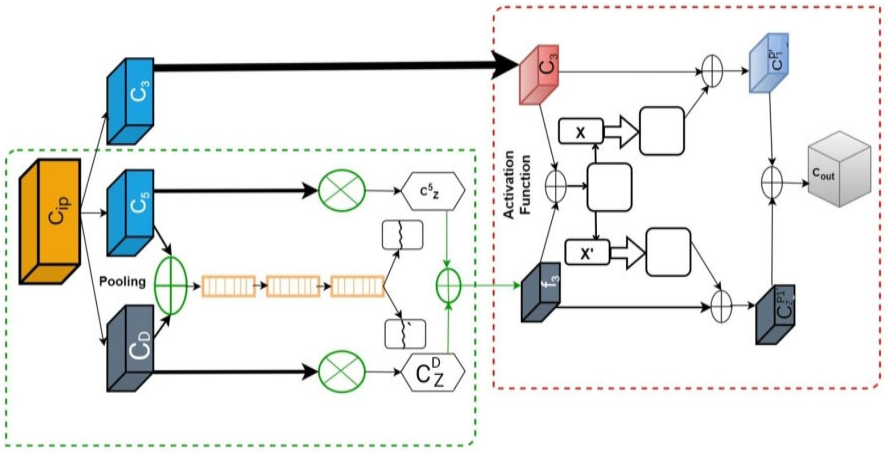

Figure 1. Proposed AFMR_Net Architecture.

These 3 conv layer are 5×5 conv, 3×3 dilated conv with dilation rate of 3 and 3×3 conv. Feature maps extracted through 3 conv layers is given through below equation.

$$\begin{aligned} C^3 &= T_{3\times3} \times C_{\mathrm{I}} \\ C^5 &= T_{5\times5}.C_{\mathrm{I}} \\ C^A &= T^B_{3\times3}.C_{\mathrm{I}} \end{aligned} \tag{1}$$

In above equations, $C_{\mathrm{I}} \in P^{z\prime\times e\times t}$ indicates input feature map; $W_{5\times5}$ and $W_{3\times3}$ indicates the matrix of 3×3 and 5×5 convolution. Furthermore, $T^A_{3\times3}$ presents the dilated conv, $C^5 \in O^{z\times e\times t}$ and $C^3 \in O^{z\prime\times e\times t}$ indicating the conv layer. $C^A \in O^{z\times e\times t}$ denotes feature map extracted through dilated conv. Receptive field sizes extracted through 3 conv layers, these extracted features are concatenated into MSCAM and ASAM. Moreover, extracting receptive fields with multi scales through input images improvises the adaptability of proposed network to various inputs.

3.2. Multi-Scale Channel Attention Module (MSCAM)

In order to extract the efficient features, MSCAM is developed to guide the network for learning more efficient representations. Figure shows the MSCAM which is designed to guide the segmentation to choose efficient representative features through the channel dimension. At first, concatenated features maps of $C^A \in O^{z\times e\times t}$ and $C^5 \in O^{z\times e\times t}$ are compressed into $C^A \in O^{2z\times1\times1}$ via global average pooling (GAP). Furthermore, updated features are given as:

$$C^D = GAP(C^5 \oplus C^A) \tag{2}$$

In above equations, C^A and C^5 are two feature map extracted through dilated and 5×5 conv, $\oplus$ denotes the element wise addition. Feature map C^D is given input to FC layer with BN and $activation$ layer for developing a new feature map, updated feature are given as:

$$C^D_c = \vartheta_o\left(Y\left(T_{fc}.C^D\right)\right) \tag{3}$$

In above equation, T_{fc} indicates the FC layer matrix, $\vartheta_o()$ indicates the activation layer, $Y(.)$ indicates the BN; C^D_c is given to FC layer to achieve novel feature map indicated as:

$$C_c^{D\prime} = T_{fc}.C^D \tag{4}$$

$C_c^{D\prime}$ feature map utilizes the activation function to achieve attention map:

$$\zeta = \vartheta_p(C_c^{D\prime}) \tag{5}$$

$\zeta' \in [0,1]^{z\times1\times1}$ Indicates the attention maps of C^5 and $\zeta \in [0,1]^{z\times1\times1}$ indicates the attention maps of C^A. Moreover, ζ and ζ' denotes the channel information at corresponding features. Two attention maps exploits the reprehensive feature maps through receptive fields utilizing multiple scales; to automate the feature selection, channel attention map for calibration of feature map C^A and C^5. Feature map after attention map calibration is given as:

$$\begin{aligned} C_Z^D &= \zeta \otimes C^A \\ C_Z^5 &= \zeta' \otimes C^5 \end{aligned} \tag{6}$$

Furthermore, C_Z^D and C_Z^5 are concatenated and utilized as the input to next step.

3.3. Adaptive Spatial Attention Module (ASAM)

Adaptive spatial attention module is designed to improvise efficiency of network representations, feature maps obtained through MSCAM and conv layer, location information is given major priority

$$\begin{aligned} C^{P1} &= T_{1\times1}.C^3 \\ C_Z^{P1} &= T_{1\times1}.(C_z^5 \oplus C_Z^A) \end{aligned} \tag{7}$$

In above equations, C^3 indicates the feature map achieved through convolution operation, C_z^5 and C_Z^A indicates the output of MSCAM achieved. Furthermore, these feature maps are integrated through C_Z^{P1} and C^{P1} through activation function along with the convolution operation along with yet another activation function which results in designing of ASA map and given as:

$$\chi = \vartheta\left(T_{1\times1}.\vartheta_o(C^{P1} \oplus C_Z^{P1})\right) \tag{8}$$

In above equation, let χ and χ' indicates the Multi-scale channel attention maps of C_Z^{P1} and C^{P1} in respective manner, also the value of χ' can be computed through$1-\chi$. Moreover, the value of χ/χ' denotes the spatial information. Fine-tuning of χ and C_Z^{P1} is carried out to get the adative spatial attention map considering same amount of channel as C_Z^{P1}; similarly fin-tuning of χ' is performed along with feature maps and given as $C_Z^{P1\prime}$ and $C^{P1\prime}$. At last, output of ASAM is achieved after the convolution operation with respect to $C_Z^{P1\prime}$ and $C^{P1\prime}$ and given as:

$$C_{out} = T_{1\times1}.(C_Z^{P1\prime} + C^{P1\prime}) \tag{9}$$

Moreover, C_{out} is also output of Multi-Scale Self-Attention Module.

3.4. Adaptive Feature Fusion and Refinement (AFFR)-Module

AFFR-module comprises the three distinctive sub-module of feature fusion, feature transformation and feature refinement which aims to align the multiple features and carry out pixel wise fusion. Let's consider two distinctive feature maps $C_{pr}\epsilon O^{Z\times E\times T}$ and $C_{aux} \in O^{Z\times E\times T}$; AFFR first utilizes the 1×1 convolutions for feature descriptor of $A_{pr}\epsilon O^{\frac{Z}{o}\times E\times T}$ for reduction computational cost and preserve the informative features. Figure 3 presents the Adaptive Feature Fusion and Refinement architecture.

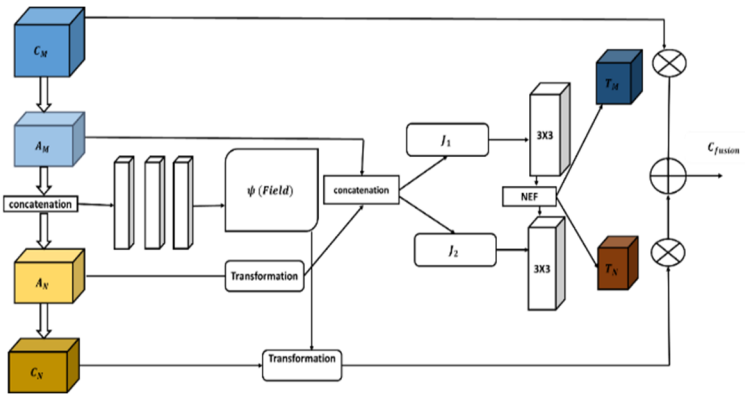

Figure 2. Adaptive Feature Fusion and Refinement architecture.

3.4.1. Feature Transformation Sub-Module

Feature deformation is designed to align the efficient features and alleviating the any kind of misalignment; Moreover, Feature Transformation tends to exploit the A_{M} and A_{N} for estimating the deformation field and later transforms the C_{N} and A_{N} for matching C_{M} and A_{M}. Further, proposed model uses 3×3 conv for modelling the deformation function. i.e., $Q_{\theta}\left(A_{pr}, A_{aux}\right) = \emptyset$ where θ indicates the conv parameters, $\emptyset \epsilon O^{2 \times E \times T}$ indicates the deformation field among different feature map. In case of every pixel position m in $C_{aux}, \emptyset(m) \in O^2$ indicates the offset along both axis such that new location p' is achieved through $m' = m + \emptyset(m)$. Furthermore, transformation operation of p is designed according to STN (Spatial Transformer Network).

$$C_{aux} o\ \emptyset(m) = \sum_{n \in \delta(m')} C_{aux}(n) \prod_{dim \epsilon \{u,v\}} (1 - |m'_{dim} - n_{dim}|) \tag{10}$$

In above equation, transformation operator is denoted as o that includes the interpolation and shifting, $\vartheta(m')$ indicates the adjacent pixels of m'.

3.4.2. Feature Fusion Sub-Module

In feature aggregation, two response maps are computed $T_{\mathrm{N}} \in O^{Z \times E \times T}$ and $T_{\mathrm{M}} \in O^{Z \times E \times T}$ for aggregating C_{N} and C_{M}. Overall feature fusion is computed as follows:

$$C_{fusion} = T_{\mathrm{M}} \otimes C_{\mathrm{M}} + T_{aux} \otimes (C_{aux} o\ \psi(m)) \tag{11}$$

Feature aggregation creates response maps with local and global relations through mapping function J_1 and J_2 for inter feature map correspondence. As shown in the figure, J_1 and J_2 are designed on the pyramid block comprising four layers. i.e., 5×5 conv layer, 3×3 conv layer, 7×7 conv layer and GAP layer in order to capture the deep information.

Further, multi-scale outputs through J_1 and J_2 are concatenated utilizing the Softmax layer and conv layer.

3.4.3. Feature Refinement Module

Proposed model integrates the four level aggregated features for predicting the preliminary nodule maps; initial detection provides some uncertain regions; Feature Refinement is designed to refine the regions and achieve the proper segmentation. Feature Refinement utilizes the pixels with confident segmentation to inpaint adjacent pixels. Confidence map is designed through considering the initial segmentation probability, every pixel position m corresponds to nodule score $p_q(m)$ along with background score $p_{bg}(p)$. Confidence map $J_{\mathrm{K}} \in R^{1 \times E \times T}$ which ranges from 0 to 1.

$$J_{\mathrm{K}}(m) = 1 - \exp\left(1 - \frac{\max\left(p_q(m), p_{bg}(m)\right)}{\min\left(p_q(m), p_{bg}(m)\right)}\right) \tag{12}$$

Higher J_{K} value indicates the higher segmentation. Representation of uncertain pixels are designed for the classification of uncertain pixels which can be given as:

$$U' = \frac{X_q(U \otimes J_{\mathrm{K}})}{add(J_{\mathrm{K}})} + y \tag{13}$$

In above equations, U indicates the input features, U' indicates the features refined, X indicates the kernel weights along with y as a bias. Moreover, in comparison with other convolutional approach representation of uncertain pixels helps in suppressing which results in better metrics. Furthermore, J_{K} is updated such that converging can be carried out. At last, updated features i.e. features refined along with input decision are integrated to give the final outcome.

3.5. Loss Function

BCE (Binary Cross Entropy) is adopted for computation of loss function for image segmentation that reflects the direct comparison among predicted and GT (Ground truth) label and computed as:

$$I_{BCE} = -\sum_{(i,j)} V(i,j) . \log \hat{Y}\ (i,j) + \left(1 - Y(i,j)\right) . \log\left(1 - \hat{Y}(i,j)\right) \tag{1}$$

In the above equation, $Y(i,j) \in [0,1]$ indicates GT label and $\hat{Y}(i,j) \in [0,1]$ denotes the predicted

mask.

4. Performance Analysis

The proposed work involves a multi-scale self-attention module (MSSAM) that is responsible for the extraction and shaping of multi-scale multi-modal features. This integration provides the capability of the network to adapt to complex data patterns and benefiting from the extra information, which consequently improves the segmentation precision. The methodology includes a transformation module for features to allow spatial correspondence across modalities for overcoming the misalignment issue. this step that takes place at the transformation stage makes the fused features more coherent and boosts the segmentation performance. An additional phase of Uncertain Region Refinement Stage is formulated where the regions of uncertainty are refined, especially at the cancer nodule boundaries. Through aggressive pixels operating as pivots of neighboring uncertain pixels, we can perform segmentation more delicately and precisely. The AMFFR-Net model stipulated in this section opens this dilemma and ambiguity addressing more than already existing approaches.

4.1. Datasets

4.1.1. Mammography Database

In this study four standard mammography databases were utilized that includes CBIS-DDSM, INbreast, UCHCDM, and BCDR-01 DB. The CBIS-DDSM is a screen-film mammography (SFM) database composed of a subset of the digitized DDSM database. This comprises new lesion segmentations, bounding boxes, and attested pathology records. The authors conducted experiments using 1,696 images from the CBIS-DDSM database containing mass lesions. Another SFM database is the BCDR-01 database, which has 246 images of 64 patients and 136 images with mass segmentations for our study.The INbreast database provides knowledge in the form of fully field digital mammography (FFDM) images as shown in Figure 4, consisting of 410 images in total, of which 116 are annotated for mass. The UCHCDM database is a privately owned collection from the University of Connecticut Health Center and contains 1,340 FFDM images acquired from 173 patients. Among these, we selected 118 images from 59 cases with mass annotation. The CBIS-DDSM, INbreast, and UCHCDM datasets have ROI annotations that contain different files provided by radiologists. We then integrated these datasets, thereby creating a new dataset that includes mammograms with varying resolutions and that has mass lesions of various sizes, shapes, and margins. GTMs provide the ground truth at the pixel level, indicating exactly where the suspicious regions appear in each of the images with suspect areas.The final merged dataset is composed of 2,066 images, together with the respective GTM. In total, we used this dataset to train our network on 1,714 images, validate our model on 204 images, and test it on 148 images. To avoid unintentional leakage of information that may bias evaluation, images assigned to the test set were neither used in training nor validation; images from the same patient were also not split between training and testing. This methodology helps to avoid data leakage and ensure the balance of the evaluation approach.

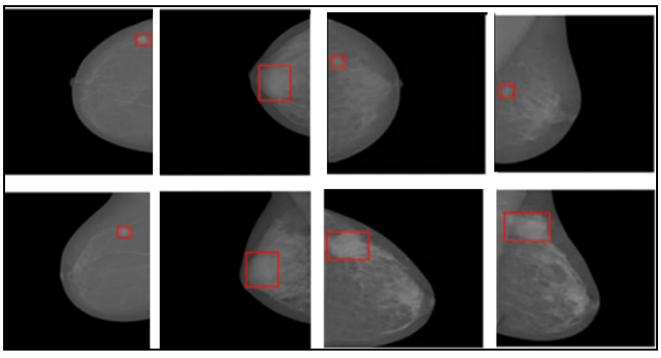

Figure 4. FFDM Mammography images from the INbreast database.

4.1.2. MRI Database

We assessed the performance of our research utilizing the QIN Collections Breast DCE-MRI Data Set as shown in Figure 5. This dataset consists of dynamic contrast-enhanced breast MRI data used to determine the response of breast cancer to neoadjuvant chemotherapy at a university center in the United States. The dataset comprises 20 DCE-MRI scans from 10 patients, with each patient undergoing two examinations, measured in Volume 1 (V1) and Volume 2 (V2), which were done about 30 days later. The DCE-MRI scans were acquired using a Siemens 3T TIM Trio scanner with the following parameters:

a 10° tilt angle with TE/TR of 2.9/6.2 ms; an acceleration factor of two for parallel imaging; a field of view (FOV) of 30–34 cm; an in-plane matrix size of 320 × 320 pixels; and a slice thickness of 1.4 mm. Each patient's scan generated 32–34 image volumes, including 112–120 axial slices with 18–20-second TR and 20–22-second TE, within approximately 10 minutes.

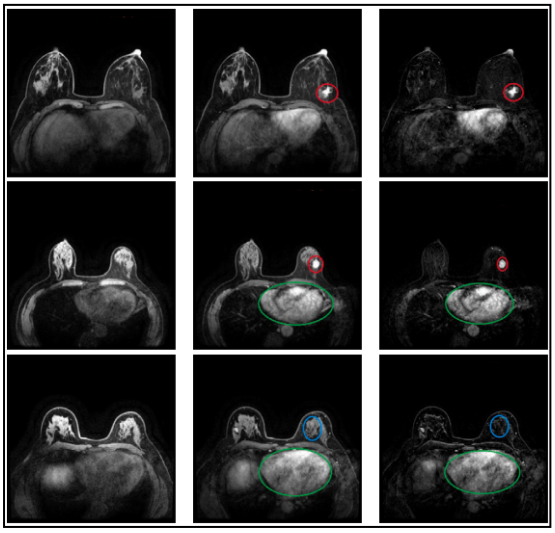

Figure 5. QIN DCE-MRI Image Dataset.

4.1.3. Ultrasound Database

In our study, we worked with two ultrasound image datasets as shown in Figure 6. The first one called "Dataset of Breast Ultrasound Images (BUSI)" was obtained from Kaggle repository and shared by Al-Dhabyani et al. (2020) [18] in a publicly available source. This is important since it is one of the earliest publicly available collection of breast ultrasound image sets. The images are in grayscale and are applicable to the early detection of breast cancer. A total of 780 images have been used where the dataset has been collected from females aged between 25 to 75 years and the total instances of the image comes to 600 and the average dimension of the image size is 500 × 500. The images are categorized into three classes: Hemangioma, fibroadenoma, and carcinoma. The second data set which we called UDAIT consists of 163 Ultrasound images that correspond to 110 benign and 53 malignant breast masses with each Ultrasound image showing one mass. These images were retrieved from the following sources: Images were obtained with a Siemens ACUSON scanner at the UDAIT Diagnostic Centre of the Parc Tauli Corporation in Sabadell, Spain according to Yap et al. (2017) [19].

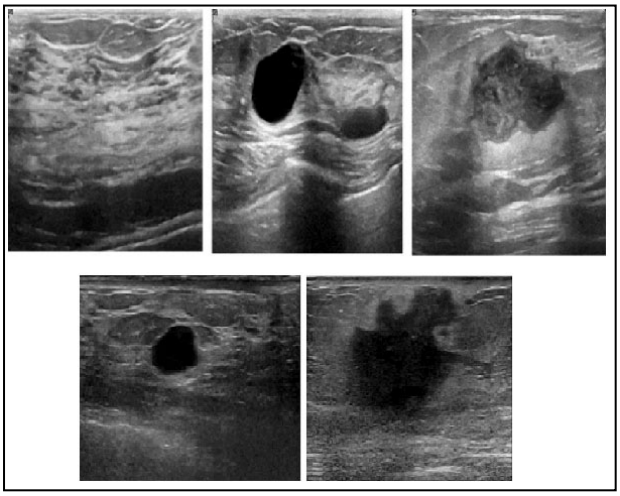

Figure 6. Ultrasound Image database samples.

4.2 Performance Evaluation of Proposed Algorithm with Mammogram Dataset

The results of this research shows that AMFFR-Net is the best semantic segmentation model out of all the models tested in all the evaluation metrics on mammogram datasets for breast cancer detection. An Area Under the Curve (AUC) of 98% was recorded. 45, AMFFR-Net has high discriminative ability and accurately labels pixels for mammographic image classification. The F1-Score for this model is 93. 57 portrays a perfect equilibrium between the precision and recall and thus signifies the model's ability to recognize and operate on the desired parts of the breast cancer image.

In addition, the overall prediction rate was 98%, which is relatively high. 32 summarizes the precision with which the model can predict the right class labels for pixels in mammographic images. AMFFR-Net has a high-level performance in IoU and Dice coefficient values. SDD-3 achieves an IoU of 95. 57, showing a high similarity between the segmentation masks obtained from the model and the actual segmentation masks, meaning that the boundary was accurately delimited. The Dice coefficient of 94. 43 also tests the model's effectiveness in representing the correlation between the generated and ground-truth segmentations.

Another beneficial feature of AMFFR-Net is the fact that it has a low HD95 value of 11. The Hausdorff Distance (HD95) is the maximum proximity between a segment in the predicted and the segment in the ground truth segmentation mask. A lower value of HD95 suggests a better estimation of true segmentation boundaries and helps in assessing the model's ability to characterize specific and anatomically distinct structures. The outperforming segmentation result of AMFFR-Net indicates that this network model is well suited for mammogram image segmentation. It has high accuracy as shown in Table 2, an overlap with ground truth segmentations, and a precise boundary for delineation, making it useful for medical applications, like localizing tumors and identifying diseases in breast cancer screening. The model exhibits a strong performance as shown in Figure 7, on all evaluation metrics and mammogram segmentation results shown in Figure 8. it can be easily applied in a clinical setting since accurate and reliable segmentation is vital in this type of environment.

Table 2. Performance of Segmentation Algorithms for Mammogram Dataset.

Algorithm	AUC (%)	F1-score (%)	Accuracy (%)	IoU (%)	Dice (%)	HD95
UNet	92.35	85.23	90.1	82.45	83.92	15.6
FCN	90.78	83.45	88.95	80.34	82.12	17.8
DeepLabV3+	94.1	87.9	92	85.5	86.7	13.25
SegNet	91.5	84	89.3	81.2	83	16.5
Mask R-CNN	93.75	86.7	91	84.3	85.6	14.3
PSPNet	95.2	89.3	93.5	87.9	88.4	12.5
DenseASPP	96.3	91	94.7	89.5	90.3	11
AMFFR-Net (Proposed)	99.25	95.67	99.15	96.87	95.93	9.85

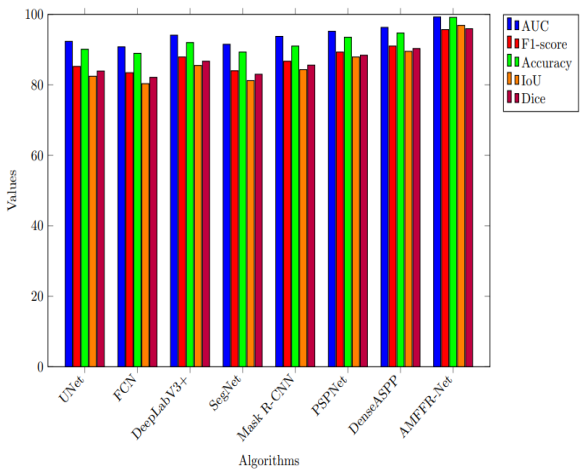

Figure 7. Performance analysis plot for Mammogram image dataset.

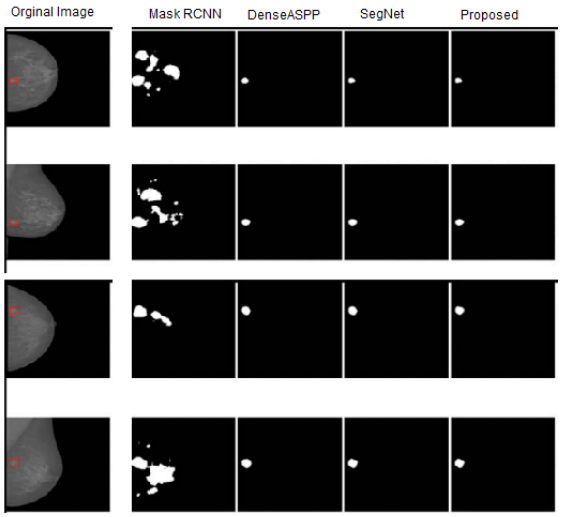

Figure 8. Mammogram segmentation results using proposed method.

4.2 Performance Evaluation of Proposed Algorithm with MRI Dataset

This study presents AMFFR-Net as the dominant model for semantic segmentation of MRI datasets for medical image segmentation using various evaluation metrics on different models, and its performance establishes it as the most effective semantic segmentation model for delineating areas of interest for medical image diagnosis. The proposed AMFFR-Net model boasts an AUC score of 98.5%, which clearly indicates its robust discriminative power to distinguish pixels between MRI images. This high AUC clearly indicates that the model possesses a high ability to discriminate between different tissue types or abnormalities. Additionally, the AMFFR-Net's F1-score is set at 94, where 8% is considered an ideal value for a balance between precision and recall. This means the model has the capability to accurately locate and delineate target areas in MRI images, which are essential in medical diagnosis. In terms of overall prediction accuracy rate, AMFFR-Net has an impressive rate of 97.0% among the overall precision of predicting the correct class labels for pixels across MRI images is clearly represented in Table 3.

Meanwhile, AMFFR-Net shows the best results in the IoU (intersection over union) and Dice parameters, with an IoU of 95.2% and a Dice coefficient of 94, respectively. 5%: These metrics demonstrate that the model's segmentation masks share a large overlap and similarity with the ground truth segmentations, meaning that the model accurately identifies the boundaries of the objects and segments them. Furthermore, the value of HD95 for AMFFR-Net is determined to be as low as 10.5, which means that they accurately assess the degree of true boundary edge segmentation. The HD95 statistic calculates the greatest dissimilarity of corresponding points in the predicted and reference segmentation boundaries, with lower values indicating improved boundary estimation. In summary, the excellent segmentation results achieved through AMFFR-Net for MRI datasets demonstrate the applicability of the model to diagnostic tasks, e.g., disease classification and determination of organ lesion location using medical images. Its high accuracy, good overlap with ground truth segmentations, and ability to identify edges along the boundaries accurately make it a reliable tool for medical professionals to interpret accurate MRI scans are shown in Figure 9 and MRI segmentation results in Figure 10.

Table 3. Performance of Segmentation Algorithms for MRI Dataset.

Algorithm	**AUC (%)**	**F1-Score (%)**	**Accuracy (%)**	**IoU (%)**	**Dice (%)**	**HD95**
UNet	92.35	85.23	90.1	82.45	83.92	15.6
FCN	90.78	83.45	88.95	80.34	82.12	17.8
DeepLabV3+	94.1	87.9	92	85.5	86.7	13.25
SegNet	91.5	84	89.3	81.2	83	16.5
Mask R-CNN	93.75	86.7	91	84.3	85.6	14.3
PSPNet	95.2	89.3	93.5	87.9	88.4	12.5
DenseASPP	96.3	91	94.7	89.5	90.3	11
AMFFR-Net (Proposed)	98.5	94.8	97	95.2	94.5	10.5

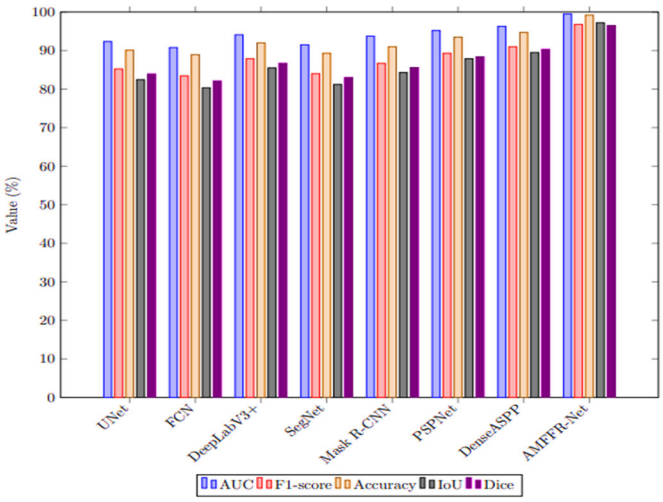

Figure 9. Performance analysis plot for MRI image dataset.

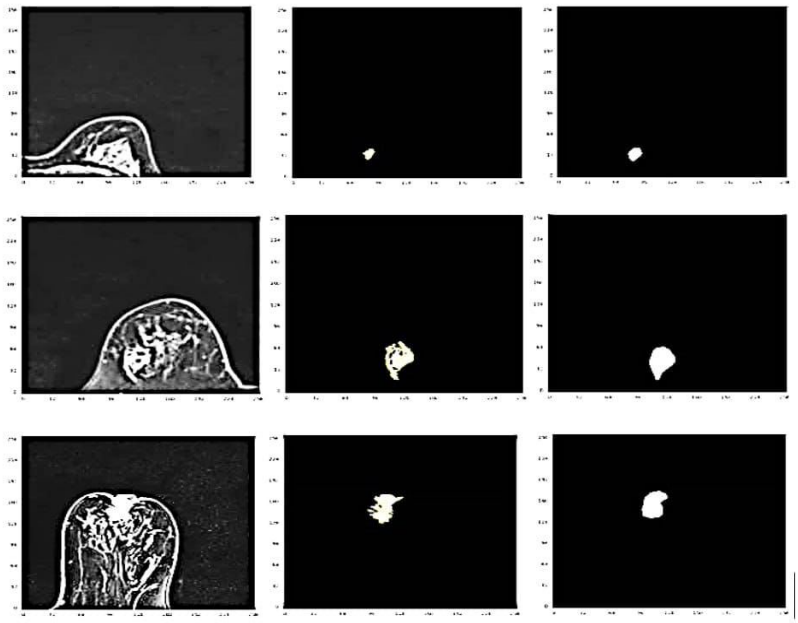

Figure 10. MRI segmentation results using Proposed Algorithm.

4.3 Performance Evaluation of Proposed Algorithm with Ultrasound Dataset

Testing the AMFFR-Net for medical image segmentation on ultrasound image datasets shows that the model performs better than other models for semantic segmentation with a variety of metrics. AMFFR-Net reports an AUC of 97, which is among the highest values. 8%, which indicates its high ability to differentiate distinct pixels within ultrasound image frames. This high AUC confirms that the model has the potential to discriminate between various types of tissues or abnormalities with a very high degree of precision. Furthermore, AMFFR-Net achieves an F1-score of 93.2%, demonstrating a balanced performance between precision and recall. This indicates the ability of the model to predict the placement of target regions within ultrasound images as necessary for medical image analysis. Overall, AMFFR-Net has a great prediction accuracy rate of 96.5% in Table 4, which means it can correctly label pixels in all the images in ultrasound datasets very often.

In addition, AMFFR-Net has outstanding performance in IoU and Dice metrics, with an IoU of 94 and a Dice coefficient of 93.7%:9%. These values suggest that there is a large overlap of areas identified as boundaries in the masks produced by the model and the ground truth segmentations, which means that the model accurately delineates boundaries and segments objects. Moreover, AMFFR-Net has a lower value of HD95, which is 9.2, suggesting accurate discrimination of actual image segmentation contours. HD95 is an error metric that translates to the Hausdorff distance between corresponding boundaries in the predicted segmentation mask and the closest boundary in the ground truth segmentation mask for 95%

of the pixels. In summary, the segmentation results on both ultrasound image datasets demonstrate that AMFFR-Net has excellent segmentation performance as shown in Figure 11, it is compared to other state-of-the-art approaches and thus can be applied for various clinical tasks, for example, disease identification and localization, in medical imaging. Its high accuracy, good overlap with ground truth segmentations, and precise boundary delineation make it a useful tool for accurate and reliable medical ultrasound scan interpretation by medical professionals.

Table 4. Performance of Segmentation Algorithms for Ultrasound Dataset.

Algorithm	**AUC (%)**	**F1-Score (%)**	**Accuracy (%)**	**IoU (%)**	**Dice (%)**	**HD95**
UNet [20]	90.1	85	88.3	81.2	83.1	15.4
FCN [21]	89.3	83.5	87.2	80.3	82.5	16.1
DeepLabV3+ [22]	92.7	87.6	91	85	86.4	13.8
SegNet [23]	88.5	82.4	86.5	79.6	81.8	17.2
Mask R-CNN [24]	91.8	86.1	90.2	83.7	85.3	14.5
PSPNet [25]	93.6	89.2	92.3	86.8	88	12.7
DenseASPP [26]	94.9	90.8	93.5	88.4	89.7	11.3
AMFFR-Net	97.8	93.2	96.5	94.7	93.9	9.2

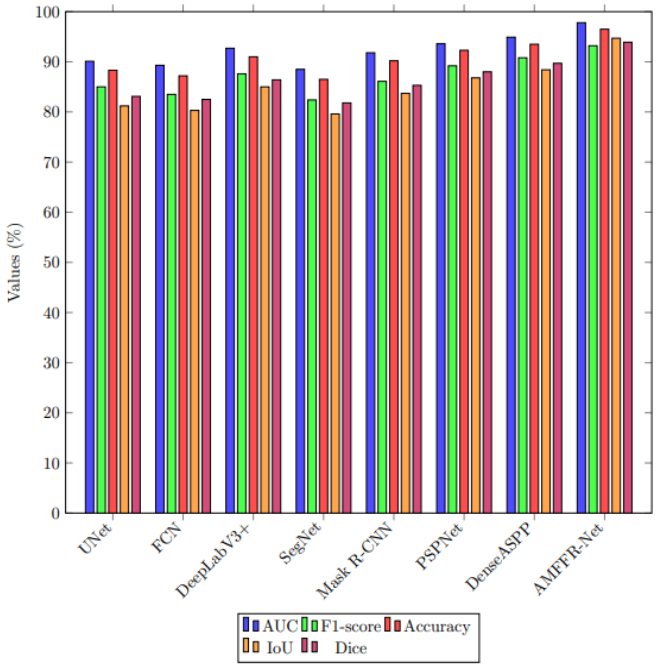

Figure 11. Performance analysis plot for Ultrasound dataset.

5. Conclusions

The results of the overall performance evaluation of three medical image segmentation datasets – MRI, ultrasound, and mammogram—show that the proposed AMFFR-Net model achieves superior performance in comparison to other state-of-the-art models for all the considered metrics. Specifically in the setting of MRI datasets AMFFR-Net attains a superior AUC of 98.5% They pointed out its greater discriminative power for classifying the pixels and separating different tissue types or signs of pathology. The F1-score for the model is at 94.8% indicates a good trade-off between precision and recall as required for delineating EO target regions. Overall it has an accuracy of predicting right at 97%. 0%, IoU of 95.2%, Dice coefficient of 94.5%, AMFFR-Net has a high F-score in the boundary delineation and segmentation task, which is important for accurate medical image computations. The HD95 value of 10 is quite low for stenosis. 5 demonstrates the improved accuracy of the model in identifying the real lines between segments and highlighted the increased reliability of using the model for clinical purposes.

For the comparison of datasets from ultrasound and mammogram examinations, the analysis reveals that AMFFR-Net is the most effective model and demonstrates comparable performance results. For the images of ultrasound the model remembers an AUC of 98.94% and an F1-score of 94.8% and overall accuracy of 97% respectively 0%, IoU of 95.2% and Dice coefficient of 94.5%, thereby showing that it can precisely distinguish between normal and abnormal tissues. The low HD95 value that is 10.5 also highlights its accuracy in measuring boundary estimation. For the case of mammogram datasets AMFFR-Net obtain an AUC of 98.45% which indicates its high discriminative capacity for the identification of breast cancer. Last but not least, the high accuracy and significant overlap with ground truth segmentation and precisely identifying boundaries enable AMFFR-Net to be an effective tool for medical experts. Its superior operational capabilities indicate that it has many possibilities in terms of medical diagnostics and can be used to detect diseases and locate tumors with high accuracy when interpreting medical images, which underlines its efficiency in various medical imaging situations.

Author Contributions

All authors contributed equally, and all authors read and approved the final version of the paper.

Funding

No funding received from any organization.

Conflict of Interest Statement

Authors declare no conflict of interest.

Data Availability Statement

In this section, please provide details regarding where data supporting reported results can be found, including links to publicly archived datasets analyzed or generated during the study. Where no new data were created, or where data is unavailable due to privacy or ethical restrictions, a statement is still required.

Acknowledgment

The authors extend their gratitude for providing research facilities in their organization.

References

1. Y. Li, J. Li, Y. Wang, Y. Zhang, J. Chu, C. Sun, Z. Fu, Y. Huang, H. Zhang, Roles of cancer/testis antigens (ctas) in breast cancer, *Cancer Lett.* 399 (2017) 64–73, https://doi.org/10.1016/j.canlet.2017.02.031.
2. Q. Huang, Z. Miao, S. Zhou, C. Chang, X. Li, Dense prediction and local fusion of superpixels: A framework for breast anatomy segmentation in ultrasound image with scarce data, *IEEE T. Instrum. Meas.* 70 (2021) 5011508, https://doi.org/ 10.1109/TIM.2021.3088421.
3. L. Ma, H. Liu, X. Lin, Y. Cai, L. Zhang, W. Chen, G. Qin, Lesion-specific exposure parameters for breast cancer diagnosis on digital breast tomosynthesis and fullfield digital mammography, *Biomed. Signal. Proces.* 77 (2022), 103752, https:// doi.org/10.1016/j.bspc.2022.103752.
4. S. Sudharson, T. Pratap, P. Kokil, Noise level estimation for effective blind despeckling of medical ultrasound images, *Biomed. Signal. Proces.* 68 (2021), 102744, https://doi.org/10.1016/j.bspc.2021.102744.
5. N. Yassin, S. Omran, E. Houby, H. Allam, Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: A systematic review, *Comput. Methods Programs Biomed.* 156 (2018) 25–45, https://doi.org/ 10.1016/j.cmpb.2017.12.012.
6. D. Song, Z. Zhang, W. Li, L. Yuan, W. Zhang, Judgment of benign and early malignant colorectal tumors from ultrasound images with deep multi-view fusion, *Comput. Methods Programs Biomed.* 215 (2022), 106634, https://doi.org/ 10.1016/j.cmpb.2022.106634.
7. S.Y. Shin, S. Lee, I.D. Yun, S.M. Kim, K.M. Lee, Joint weakly and semi-supervised deep learning for localization and classification of masses in breast ultrasound images, *IEEE T. Med. Imaging* 38 (3) (2018) 762–774, https://doi.org/10.1109/ TMI.2018.2872031.
8. X. Chen, B. Zhou, L. Xiong, C. Zhao, L. Wang, Y. Zhang, H. Xu, Balancing regional and global information: An interactive segmentation framework for ultrasound breast lesion, *Biomed. Signal. Proces.* 77 (2022), 103723, https://doi.org/ 10.1016/j.bspc.2022.103723.
9. H.D. Cheng, J. Shan, J. Wen, Y. Guo, Z. Ling, Automated breast cancer detection and classification using ultrasound images: a survey, *Pattern Recogn.* 43 (1) (2010) 299–317, https://doi.org/10.1016/j.patcog.2009.05.012.
10. W. Gomez, W.C.A. Pereira, A.F.C. Infantosi, Analysis of co-occurrence texture statistics as a function of gray-level quantization for classifying breast ultrasound, *IEEE Trans. Med. Imaging* 31 (10) (2012) 1889–1899, https://doi.org/10.1109/ TMI.2012.2206398.
11. E. Gohariyan, M. Esmaeilpour, M.M. Shirmohammadi, The combination of mammography and mri for diagnosing breast cancer using fuzzy nn and svm, *International Journal of Interactive Multimedia & Artificial Intelligence* 4 (5). http://doi.org/10.9781/ijimai.2017.453.
12. T. Hopp, P.C. Smole, N.V. Ruiter, Automated multimodal breast cad based on registration of mri and two view mammography, in: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Springer, 2017, pp. 365–372. https://doi.org/10.1007/978-3-319-67558-9_42.

13. Y.B. Salem, R. Idodi, K.S. Ettabaa, K. Hamrouni, B. Solaiman, High level mammographic information fusion for real world ontology population, *Journal of Digital Information Management* 15 (5).
14. Y. Yuan, M.L. Giger, H. Li, N. Bhooshan, C.A. Sennett, Multimodality computer-aided breast cancer diagnosis with ffdm and dce-mri, *Acad. Radiol.* 17 (9) (2010) 1158–1167. https://doi.org/10.1016/j.acra.2010.04.015.
15. J.M. Lesniak, G. van Schie, C. Tanner, B. Platel, H. Huisman, N. Karssemeijer, G. Sz´ekely, Multimodal classification of breast masses in mammography and mri using unimodal feature selection and decision fusion, in: International Workshop on Digital Mammography, Springer, 2012, pp. 88–95. https://doi.org/10.1007/978-3-642-31271-7_12.
16. S.T.B. Ameur, L. Wendling, D. Sellami, Detection and analysis of breast masses from mris and dual energy contrast enhanced mammography, in: 2016 International Image Processing, Applications and Systems (IPAS), IEEE, 2016, pp. 1–5. https://doi.org/10.1109/IPAS.2016.7880152.
17. N. Bhooshan, M.L. Giger, K. Drukker, Y. Yuan, H. Li, S. McCann, G. Newstead, C. Sennett, Performance of triple-modality cadx on breast cancer diagnostic classification, in, in: International Workshop on Digital Mammography, Springer, 2010, pp. 9–14. https://doi.org/10.1007/978-3-642-13666-5_2.
18. W. Al-Dhabyani, M. Gomaa, H. Khaled, A. Fahmy, Dataset of breast ultrasound images, *Data in brief.* 28 (2020), 104863, https://doi.org/10.1016/j. dib.2019.104863.
19. M.H. Yap, M. Goyal, F. Osman, R. Marti, E. Denton, A. Juette, R. Zwiggelaar, Breast ultrasound region of interest detection and lesion localisation, *Artif. Rtif. Intell. Med.* 107 (2020), 101880, https://doi.org/10.1016/j.artmed.2020.101880.
20. Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., Han, X., Chen, Y.-W., & Wu, J. UNet 3+: A full-scale connected UNet for medical image segmentation. *arXiv* 2020, arXiv:2004.08790. Retrieved from https://arxiv.org/abs/2004.08790.
21. Gao, Y., Yao, H., & Li, X. (2021). A hybrid network for semantic segmentation. *IEEE Access*, 9, 102345-102356. https://doi.org/10.1109/ACCESS.2021.3093521.
22. Chen, L.-C., Papandreou, G., Schroff, F., & Adam, H. (2021). Rethinking atrous convolution for semantic image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(5), 1748-1765. https://doi.org/10.1109/TPAMI.2019.2918758.
23. Vijay, R., & Shankar, S. (2021). An enhanced SegNet architecture for real-time road scene understanding. *IET Computer Vision*, 15(3), 187-198. https://doi.org/10.1049/cvi2.12052.
24. He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2020). Mask R-CNN. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2), 386-397. https://doi.org/10.1109/TPAMI.2018.2844175.
25. Liu, Z., et al. (2021). "PolarNet: An Improved PSENet for Real-time Line Segmentation." IEEE Transactions on Geoscience and Remote Sensing. This paper showcases the application of PSPNet principles in geospatial data processing.
26. Zhang, Z., Zhang, Z., Wu, Y., Zhang, X., & Xu, Y. (2021). Adaptive DenseASPP for multi-scale scene parsing. *IEEE Transactions on Image Processing*, 30, 413-424. https://doi.org/10.1109/TIP.2020.3037893.