

An Enhanced Measurement of Epicardial Fat Segmentation and Severity Classification using Modified U-Net and FOA-guided XGBoost

Rajalakshmi K^{1*}, Palanivel Rajan S²

¹Department of Artificial Intelligence and Data Science, Kamaraj College of Engineering and Technology, K. Vellakulam, 625701, Tamil Nadu, India, k.rajalakshmiphd@gmail.com

²Department of Electronics and Communication Engineering, Velammal College of Engineering and Technology, Madurai, 625009, Tamil Nadu, India

Abstract: The amount of epicardial fat around the heart has a significant impact on cardiovascular function and requires precise measurement for timely treatment. In this work, an improved U-Net architecture is proposed for accurate segmentation of epicardial fat in computer tomography (CT) images. The proposed method integrates a modified squeeze-and-excitation (MSE) block and a multi-scale dense (MS-D) convolutional neural network (CNN) to improve feature extraction. In addition, a metaheuristic optimization algorithm from falcon optimization algorithm (FOA) is used for efficient feature selection. The selected features are then classified using the XGBoost algorithm to determine the fat severity. Experimental evaluations on a CT image dataset show the superior segmentation performance of the proposed U-Net compared to existing architectures. It achieves a mean intersection over union (*Mean IOU*) of 89.5 %, a mean Dice score (*MDS*) of 94.3 %, and a Pearson correlation coefficient (*PCC*) of 0.973. FOA-guided feature selection further increases the accuracy of severity classification. The overall classification accuracy of the model is 98 %. These results highlight the technological advancements and measurement accuracy of the proposed U-Net architecture. They also demonstrate the suitability of the model to improve cardiovascular risk assessment and management.

Keywords: epicardial fat segmentation, U-Net architecture, falcon optimization algorithm, feature selection, severity classification

1. INTRODUCTION

Cardiovascular diseases (CVDs) are the leading cause of death worldwide. Epicardial adipose tissue (EAT) is one of the most important risk factors for CVDs [1]-[3]. EAT is the fat deposit in the heart that is located between the myocardial surface and the visceral layer of the pericardium. The most important functions of EAT include protection against hypothermia, mechanical protection of the coronary circulation and energy supply to the myocardium. It also secretes adiponectin from the epicardial adipocytes. The volume and thickness of fat affect cardiovascular activity and is the main cause of obesity. Recently, the image processing-based EAT approach has gained more attention due to its non-invasiveness [4]. This technique assists healthcare providers in analyzing volume, density, spatial distribution, etc. These imaging techniques include various modalities such as computer tomography (CT), magnetic resonance imaging (MRI), and echocardiography for accurate quantification and visualization of EAT.

This quantification supports risk stratification and disease prognosis. Recently, the use of machine learning (ML) and

deep learning (DL) models in cardiovascular imaging has gained greater attention [5]. This model is used by clinicians to analyze complex data and extract meaningful information [6]. The well-known ML algorithms such as Ad Boost, Random Forests, and Decision Trees are used for image classification, segmentation, and feature extraction. The well-known DL algorithm such as convolutional neural networks (CNNs) has gained importance due to its ability to automatically learn features from images. Recently, popular DL architectures such as U-Net, DenseNet, and ResNet are well suited for medical image analysis such as tumor detection, organ segmentation, and disease classification. This model's ability to learn complex patterns from large datasets contributes to improved patient care.

A medical image segmentation model using a feature compression pyramid network was developed [7]. An innovative interactive segmentation framework utilizing DL principles was presented [8]. An ensemble of U-Net architectures was developed for kidney tumor segmentation [9]. A segmentation model based on a 3-D CNN called HyperDenseNet has been proposed for medical image processing [10]. An attention mechanism-based model was

introduced for medical image segmentation [11]. DSI-Net, which includes a classification section, a coarse segmentation branch, and a fine segmentation branch, was introduced [12]. A smaller model called PocketNet was proposed for the combined segmentation and classification process [13].

Transformer networks use self-attention mechanisms and are able to capture global dependencies within images. A spatio-temporal transformer network has been proposed for medical image processing [14]. Traditional pixel-wise classifiers in deep learning do not take into account the output structure and the inter-dependence. To address this, a new training approach was introduced [15]. A new deep learning model was developed for the epicardial fat segmentation and classification in non-contrast CT images [16]. The dual U-Net concept was proposed for epicardial fat detection [17].

In light of these advancements, this work aims to contribute to the evolving model of cardiovascular research by proposing a novel approach for accurate EAT assessment and severity classification. The proposed model integrates DL models with architectural modifications and optimization frameworks for assessing EAT severity. The contributions of the proposed work are as follows:

- Performing a modified U-Net based segmentation to obtain strong features;
- Applying an optimization algorithm to select the best features;
- Developing an XGBoost classification model for EAT severity classification;
- The proposed approach is compared with other models.

The rest of the work is organized as follows: Section 2 discusses the proposed model for EAT segmentation and classification. This is followed by a results and discussion section in Section 3. Section 4 discusses a conclusion.

2. PROPOSED MODEL

This section first describes the proposed U-Net model for EAT segmentation and XGBoost model-based severity classification. First, the EAT regions of the image are segmented using the proposed modified squeeze-and-excitation gated recurrent unit U-Net (MSE-GRU-U-Net). This architecture includes a MSE and multi-scale dense (MS-D) network for accurate EAT segmentation. Then the ML model of XGBoost is applied to calculate the severity based on the trained features extracted from the EAT regions. The entire workflow is shown in Fig. 1.

MSE-GRU U-Net

The proposed MSE-GRU-U-Net architecture comprises a symmetric structure with two core segments: the encoder and the decoder, as shown in Fig. 2. The encoder segment focuses on feature extraction and the decoder component is used for precise feature positioning. This architecture includes residual blocks, pooling layers, MSEs, MS-D, and 4 up-sampling blocks for the $512 \times 512 \times 1$ input images.

The residual learning strategy is applied by integrating shortcut connections into conventional U-Nets. These connections make the model suitable for deeper training and prevent degradation. Throughout the model, features undergo a series of operations including convolutions, feature extraction, and pooling. The size of the final binary segmented image is $512 \times 512 \times 1$.

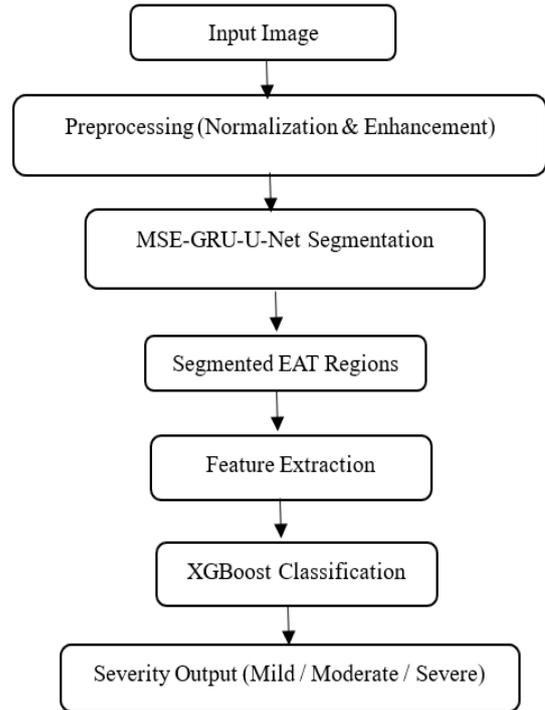


Fig. 1. Workflow of the proposed model.

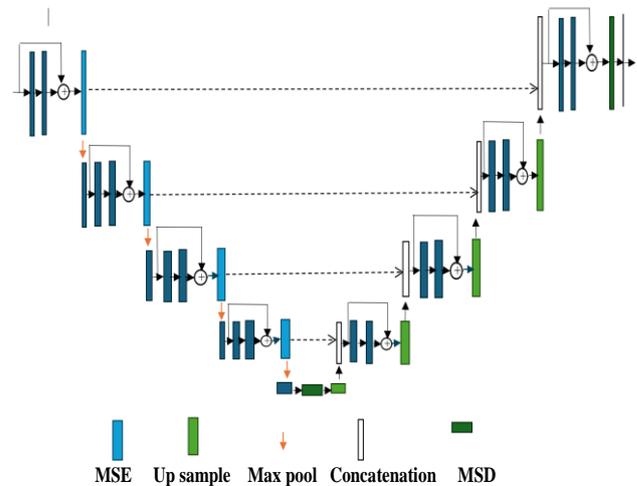


Fig. 2. The proposed model.

To solve the vanishing gradient problem, batch normalization and rectified linear unit (ReLU) activation units are used in each convolution operation. In addition, the MS integrates a bi-directional gated recurrent unit (BiGRU) layer to adaptively extract image features. In the U-Net, the BiGRU layer is used for channel attention and effective segmentation.

In addition, the MS-D network is used as a transition layer after the encoder and at the output of the decoder. The insertion of the MS-D network into the U-Net improves the model's ability to learn multi-scale contextual information [18]. It also effectively solves the resolution reduction problem caused by multiple down-sampling operations. The dilated convolutions allow the network to gather contextual information without over down-sampling the feature maps.

Modified squeeze-and-excitation

In CNN, the squeeze-and-excitation (SE) block is inserted to increase the representational power of the model. It is used to focus on relevant features and eliminate irrelevant features [19]. The SE block works by learning channel-wise feature recalibration weights to adaptively focus on essential features. This process involves two fundamental operations: squeezing and excitement. The squeezing step aggregates channel-wise information to capture global statistics and the excitement step performs feature re-weighting based on the learned parameters. In the MSE block, the additional layers are introduced to improve feature representation and the learning capabilities of the neural networks. The GRU layers are added in the conventional SE block to better capture spatial and temporal information in the images. The MSE block includes global pooling, fully connected layers, ReLU activation, BiGRU layers, and a final sigmoid activation.

Global pooling layer:

The input feature maps are globally pooled to summarize the spatial information into a single vector for each channel. The output of the pooling layer is expressed as follows:

$$Y = \text{GlobalPooling}(X) \quad (1)$$

Fully connected layer 1:

The pooled features are passed through a fully connected layer *FC1* that facilitates transformation to a higher-level feature representation.

$$Z1 = \text{FC1}(Y) \quad (2)$$

ReLU activation:

Following the fully connected layer, a *ReLU* activation function introduces non-linearity into the network.

$$A1 = \text{ReLU}(Z1) \quad (3)$$

BiGRU layer 1:

The ReLU-activated features are passed through the first BiGRU layer, *BiGRU1*, which allows the network to capture sequential information bidirectionally.

$$H1 = \text{BiGRU1}(A1) \quad (4)$$

BiGRU layer 2:

The output of the first BiGRU layer is further processed by a second BiGRU layer, *BiGRU2* to capture more complicated sequential patterns.

$$H2 = \text{BiGRU2}(H1) \quad (5)$$

Fully connected layer 2:

Another fully connected layer *FC2* is used to perform additional feature transformations.

$$Z2 = \text{FC2}(H2) \quad (6)$$

Sigmoid activation:

Finally, a sigmoid activation function is applied to generate the output probabilities by considering the important features and suppressing the less relevant ones.

$$O = \text{Sigmoid}(Z2) \quad (7)$$

where *Y* represents the global pooled features, *Z1* represents the output of the first fully connected layer, *A1* represents the output after ReLU activation, *H1* and *H2* represent the outputs of the first and second BiGRU layers, respectively, *Z2* represents the features after the last fully connected layer, and *O* represents the output after sigmoid activation, emphasizing important features based on the learned weights.

First, the input feature maps are subjected to global pooling operations to summarize the spatial information, allowing the extraction of global statistics for each channel. The addition of multiple layers enables hierarchical feature learning, allowing the model to extract complex representations of the fat image. This hierarchical feature learning helps to capture both spatial and temporal patterns.

The excitation operation involves feature learning based on channel-wise weights. In addition, the network's adaptability allows it to dynamically focus its attention on different aspects of the pixels. The MSE block improves the standard SE by integrating multi-scale feature aggregation and frequency-aware attention. It improves contextual understanding and reduces computational effort.

MS-D network

The MS-D network architecture is an innovative approach in DL for image processing. This network combines mixed-scale dilated convolutions with dense connections to improve feature extraction and information flow between layers. This architecture comprises multiple layers of feature maps, each derived from a consistent set of operations. A sequential process of operations to generate feature maps is referred to as follows:

$$g_{ij}(\{z_0, \dots, z_{i-1}\}) = \sum_{l=0}^{i-1} \sum_{k=0}^{c_l-1} D_{h_{ijkl}, s_{ij}} z_l^k \quad (8)$$

where g_{ij} is the output feature map at position i, j derived from the previous feature maps $\{z_0, \dots, z_{i-1}\}$, $D_{h_{ijkl}, s_{ij}}$ represents the weight parameter for the dilated convolution operation, z_l^k represents the previous feature maps (z) indexed by k at position l , D denotes the total number of channels in the feature maps, and h_{ijkl}, s_{ij} denotes the parameters controlling the dilation and convolution operation. This equation represents the calculation for a specific position i, j within a feature map in the MS-D network architecture. It takes into account the input or previous feature maps, the convolutional weights, and the dilation parameters to generate the output feature map at that spatial position.

The operations include Dilated Convolutions, Pixel-wise Summation and Bias Addition. Using 3×3 pixel filters with channel-specific dilation, convolutions are applied to all previous feature maps, capturing the multi-scale information important for nuanced feature extraction. The resulting images from the convolutions are summed pixel by pixel,

enabling the integration of different spatial information across multiple scales. A constant bias is added to each pixel, which helps in introducing learnable parameters to fine-tune the feature representations. This sequential operation generates a series of feature maps to refine and extract hierarchical features from the input image.

Feature selection using falcon optimization algorithm

Feature selection is an important task to achieve higher accuracy and reduce computation time. The performance of classification mainly depends on the feature selection process. In this work, falcon optimization algorithm (FOA) is used to select the optimal features extracted from the modified U-Net architecture. The falcons represent potential solutions that navigate through the solution space. This population-based algorithm uses a set of N falcons as search agents in the challenge space. Optimal solutions are represented by the falcons' positions referred to as g_{best} . Each

falcon updates its position based on (4).

The convergence rate decreases linearly from 1 to 2 over the iterations and increases the probability of reaching a global solution. Each falcon updates its position based on only one g_{best} , which promotes a comprehensive utilization of potential solutions.

The features selected by FOA are normalized and filtered based on the mean and standard deviation to determine the best subset of features:

$$NZ_i = \frac{K_i - \mu}{S} \quad (9)$$

where K_i denotes the selected features, μ is the mean, S is the standard deviation, and NZ_i is the normalized feature vector. The final selection is determined by the comparison between the selected features and the best vector derived by FOA using a quadratic fitness function evaluated by the mean square error rate.

Pseudocode: FOA Feature Selection and XGBoost Optimization

```
# *Initialize XGBoost parameters*
xgb_params = {
    'learning_rate': 0.1,
    'max_depth': 5,
    'n_estimators': 100,
    # *Other parameters...*
}

# *Feature selection using FOA*
def FOA_feature_selection(feature_set, labels):
    # *Implement FOA for feature selection*
    # *Initialize population*
    initialize_population()
    # *Evaluate fitness of each individual*
    evaluate_fitness(feature_set, labels)
    # *Repeat iterations*
    while not stopping_criteria_met():
        # *Apply selection, crossover, and mutation*
        apply_selection_crossover_mutation()
        # *Evaluate fitness of new individuals*
        evaluate_fitness(feature_set, labels)
    # *Return selected features from the best individual*
    return best_individual_features

# *Optimize XGBoost parameters with FOA*
def FOA_optimize_xgb(xgb_params, selected_features, labels):
    best_params = xgb_params.copy()
    # *Define FOA iterations for parameter optimization*
    for iteration in range(max_iterations):
        # *Perform optimization on XGBoost parameters using FOA*
        for param in xgb_params:
            # *Set range for the parameter*
            param_range = define_parameter_range(param)
            # *Apply FOA to optimize the parameter*
            updated_param_value = optimize_parameter(param_range)
            # *Update XGBoost parameters with optimized value*
            best_params[param] = updated_param_value
        # *Train XGBoost model with updated parameters*
        trained_model = train_xgboost_model(best_params, selected_features, labels)
        # *Evaluate model performance and update best_params if needed*
        if model_performance_improved(trained_model):
            best_params = updated_params
    return best_params

# *Run feature selection and XGBoost optimization*
selected_features = FOA_feature_selection(feature_set, labels)
best_xgb_params = FOA_optimize_xgb(xgb_params, selected_features, labels)
```

XGBoost based classification

XGBoost is a ML model that works on the basis of gradient boosting algorithms. To achieve an objective function, it sequentially constructs a decision tree. The algorithm updates the tree weights to minimize the loss function by gradient descent optimization. The result of this boosting method is an ensemble of trees, whose predictions are combined using weighted averaging. The number of trees, the tree depth, the learning rate, and the regularization terms are the four most important hyperparameters of the XGBoost model.

FOA is used in this work to determine the ideal hyperparameters of the XGBoost model. To find the ideal set of hyperparameters, FOA is used to optimize a subset of these values. FOA provides effective assistance. The pseudocode for the proposed feature selection and tuning of the XGBoost model is presented above.

The first function, *FOA_feature_selection ()*, focuses on using FOA to iteratively select the most relevant features from a given set. Through population initialization, fitness evaluation, and iterative refinement, which includes updating the best positions, generating new locations, and updating the falcon locations, this process identifies the optimal set of features that are critical for EAT segmentation and severity analysis. Next, the *FOA_optimize_xgb ()* function uses FOA to optimize the parameters within the XGBoost model. It iterates through the defined parameter space and optimizes parameters such as the learning rate, max depth, and other parameters relevant to XGBoost to improve the model's efficiency in classifying the EAT severity levels. FOA increases the computational complexity due to fractional-order calculations that require memory-intensive recursive operations. It improves feature sensitivity but requires efficient GPU acceleration to maintain practical training times.

3. RESULTS AND DISCUSSION

The experimental datasets used in this work are taken from the website (<http://visual.ic.uff.br/en/cardio/ctfat/index.php>). These datasets consist of features extracted from 200 patients. The validity and repeatability of the proposed method was assessed by conducting ten-fold cross-validation experiments with the collected dataset. In each cross-validation, the dataset was divided into 70 % for training and 30 % for validation. In this dataset, segmentation faces challenges such as low contrast and artifacts. To overcome this, adaptive weighting and post-processing are performed. The performance of the proposed segmentation architecture is compared with other segmentation models in terms of mean intersection over union (*Mean IOU*), mean Dice score (*MDS*), and Pearson correlation coefficient (*PCC*).

Mean IOU evaluates the intersection between the predicted segmentation and the ground truth (GT) segmentation for multiple classes or instances. It is calculated as the intersection divided by the union of the predicted and GT masks as follows:

$$IOU = \frac{\text{Area of Overlap}}{\text{Area of Union}} \quad (10)$$

The *Mean IOU* is calculated as follows:

$$\text{Mean IOU} = \frac{1}{N} \sum_{i=1}^N \frac{TP_i}{TP_i + FP_i + FN_i} \quad (11)$$

where N is the number of classes or instances, TP_i is the true positive for class i , FP_i is the false positive for class i , FN_i is the false negative for class i .

The *Dice score* (also known as F1 score) is used to evaluate the similarity between two samples. In segmentation, it measures the overlap between the predicted and GT segmentation as follows:

$$\text{Dice score} = \frac{2 \times \text{Area of Overlap}}{\text{Area of Predicted} + \text{Area of GT}} \quad (12)$$

The *MDS* is calculated as follows:

$$\text{MDS} = \frac{1}{N} \sum_{i=1}^N \frac{2 \times TP_i}{2 \times TP_i + FP_i + FN_i} \quad (13)$$

$$\text{MDS} = \frac{1}{N} \sum_{i=1}^N \frac{2 \times TP_i}{2 \times TP_i + FP_i + FN_i}$$

The *PCC* estimates the linear correlation between two sets of data. It is used to analyze the relationship between predicted and GT pixel values. It can be calculated as follows:

$$\text{PCC} = \frac{\sum(x - x') + (y - y')}{\sqrt{\sum(x - x')^2} \times \sqrt{\sum(y - y')^2}} \quad (14)$$

where x and y are the predicted and GT values for the pixels, and x' and y' are their respective means.

The segmentation results are shown in Fig. 3. The segmentation results correspond well with the GT images. The red color represents the epicardial fat, the green color represents the mediastinal fat, and the blue color represents the gap between the epicardial and mediastinal fat. The training and validation loss of the proposed model over epochs is shown in Fig. 4.

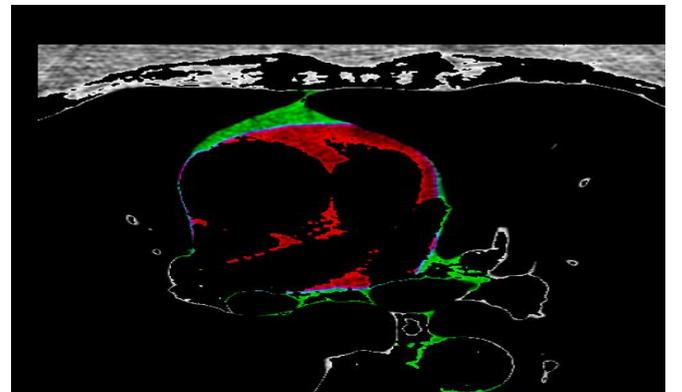


Fig. 3. Segmentation result (epicardial and mediastinal fats).

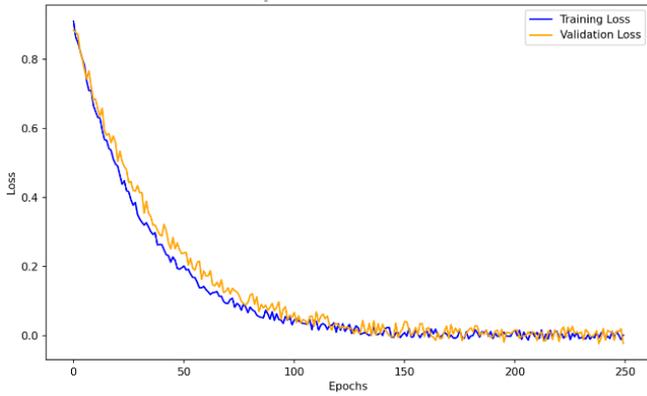


Fig. 4. Training and validation loss of the proposed model over epochs.

Table 1. Performance analysis.

Model	Mean IOU [%]	MDS [%]	PCC
FCN	68.92	73.5	0.789
U-Net	71.5	79.7	0.8721
Seg-Net	76.7	85.42	0.8905
Attention U-Net	80.2	88.54	0.932
SAR-U-Net	86.78	92.6	0.954
MSE-GRU-U-Net	89.5	94.3	0.973

The performance of the proposed and existing segmentation models is given in Table 1. The fully convolutional network (FCN) model shows a *Mean IOU* of 68.92 % and a *MDS* of 73.5 %, indicating reasonable but comparatively lower segmentation accuracy. The U-Net shows improvements with a *Mean IOU* of 71.5 % and a *MDS* of 79.7 %, indicating improved segmentation ability. Seg-Net further improves performance, achieving a *Mean IOU* of 76.7 % and a *MDS* of 85.42 %, indicating significant improvements in image region delineation. The Attention U-Net shows a *Mean IOU* of 80.2 % and a *MDS* of 88.54 %, which represents a remarkable progress in segmentation accuracy. The SAR-U-Net achieved a *Mean IOU* of 86.78 % and a *MDS* of 92.6 %, indicating exceptional precision in the accurate segmentation of image structures. The correlation coefficient between these models is between 0.789 and 0.954. The proposed MSE-GRU-U-Net models show better performance in terms of *Mean IOU*, *MDS* and *PCC* with values of 89.5, 94.3 and 0.973, respectively. The results are shown graphically in Fig. 5.

The features extracted from the segmentation model are used to train the XGBoost classifier with labels of low, high and medium severity levels. For a fair comparison, the FOA-XGBoost model is compared with the other FOA-optimized models from ID3 and naive bayes (NB). The measured values for accuracy, specificity, precision and recall are shown in Table 2. The FOA-ID3 model shows 92 % accuracy, 91 % specificity, 90 % precision, and 94 % recall rates. The FOA-NB model has 95 % accuracy, 95 % specificity, 94 % precision, and 94 % recall rates. The FOA-XGBoost model stands out with the highest performance values, achieving 97 % accuracy, 96 % specificity, 97 % precision, and 98 % recall rates demonstrating exceptional accuracy in categorizing severity levels.

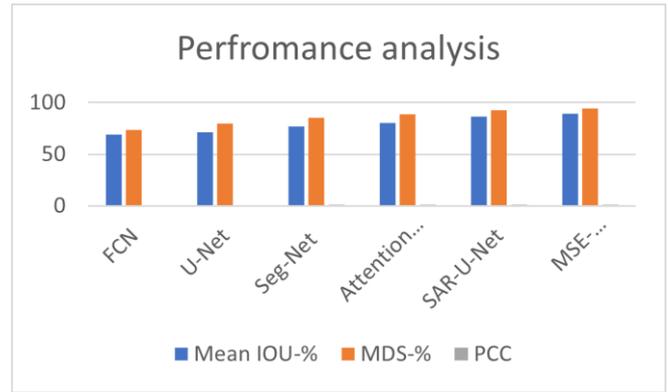


Fig. 5. Performance analysis.

Table 2. Performance analysis.

Model	FOA-ID3 [%]	FOA-NB [%]	FOA-XGBoost [%]
Accuracy	92	95	97
Specificity	91	95	96
Precision	90	94	97
Recall	94	94	98

4. CONCLUSION

In this work, a new approach to EAT segmentation is proposed. The new architecture integrates MSE with MS-Ds for accurate segmentation of fat regions. The MSE-GRU-U-Net achieves a *Mean IOU* value of 89.5 % and a *MDS* of 94.3 % for segmentation. It also has a strong correlation coefficient of 0.973, indicating a highly reliable relationship between the predicted and GT values. For classification, the FOA-XGBoost model achieves the highest accuracy of 97 %. The proposed model is fully automated and can be an accurate diagnostic tool for advanced clinical practice.

REFERENCES

- [1] Goeller, M., Achenbach, S., Marwan, M., Doris, M. K., Cadet, S., Commandeur, F., Chen, X., Slomka, P. J., Gransar, H., Cao, J. J., Wong, N. D., Albrecht, M. H., Rozanski, A., Tamarappoo, B. K., Berman, D. S., Dey, D. (2018). Epicardial adipose tissue density and volume are related to subclinical atherosclerosis, inflammation and major adverse cardiac events in asymptomatic subjects. *Journal of Cardiovascular Computed Tomography*, 12 (1), 67-73. <https://doi.org/10.1016/j.jcct.2017.11.007>
- [2] Mahabadi, A. A., Lehmann, N., Kälsch, H., Bauer, M., Dykun, I., Kara, K., Moebus, S., Jöckel, K. H., Erbel, R., Möhlenkamp, S. (2014). Association of epicardial adipose tissue and left atrial size on non-contrast CT with atrial fibrillation: The Heinz Nixdorf Recall Study. *European Heart Journal Cardiovascular Imaging*, 15 (8), 863-869. <https://doi.org/10.1093/ehjci/jeu006>
- [3] Konwerski, M., Gąsecka, A., Opolski, G., Grabowski, M., Mazurek, T. (2022). Role of epicardial adipose tissue in cardiovascular diseases: A review. *Biology*, 11 (3), 355. <https://doi.org/10.3390/biology11030355>

- [4] Davidovich, D., Gastaldelli, A., Sicari, R. (2013). Imaging cardiac fat. *European Heart Journal - Cardiovascular Imaging*, 14 (7), 625-630. <https://doi.org/10.1093/ehjci/jet045>
- [5] Greco, F., Salgado, R., Van Hecke, W., Del Buono, R., Parizel, P. M., Mallio, C. A. (2022). Epicardial and pericardial fat analysis on CT images and artificial intelligence: A literature review. *Quantitative Imaging in Medicine and Surgery*, 12 (3), 2075-2089. <https://doi.org/10.21037/qims-21-945>
- [6] Hoori, A., Hu, T., Lee, J., Al-Kindi, S., Rajagopalan, S., Wilson, D. L. (2022). Deep learning segmentation and quantification method for assessing epicardial adipose tissue in CT calcium score scans. *Scientific Reports*, 12, 2276. <https://doi.org/10.1038/s41598-022-06351-z>
- [7] Liu, Y., Zhou, J., Liu, L., Zhan, Z., Hu, Y., Fu, Y., Duan, H. (2022). FCP-Net: A feature-compression-pyramid network guided by game-theoretic interactions for medical image segmentation. *IEEE Transactions on Medical Imaging*, 41 (6), 1482-1496. <https://doi.org/10.1109/TMI.2021.3140120>
- [8] Wang, G., Li, W., Zuluaga, M. A., Pratt, R., Patel, P. A., Aertsen, M., Doel, T., David, A. L., Deprest, J., Ourselin, S., Vercauteren, T. (2018). Interactive medical image segmentation using deep learning with image-specific fine-tuning. *IEEE Transactions on Medical Imaging*, 37 (7), 1562-1573. <https://doi.org/10.1109/TMI.2018.2791721>
- [9] Causey, J., Stubblefield, J., Qualls, J., Fowler, J., Cai, L., Walker, K., Guan, Y., Huang, X. (2022). An ensemble of U-Net models for kidney tumor segmentation with CT images. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 19 (3), 1387-1392. <https://doi.org/10.1109/TCBB.2021.3085608>
- [10] Dolz, J., Gopinath, K., Yuan, J., Lombaert, H., Desrosiers, C., Ben Ayed, I. (2019). HyperDense-Net: A hyper-densely connected CNN for multi-modal image segmentation. *IEEE Transactions on Medical Imaging*, 38 (5), 1116-1126. <https://doi.org/10.1109/TMI.2018.2878669>
- [11] Ling, Y., Wang, Y., Dai, W., Yu, J., Liang, P., Kong, D. (2024). MTANet: Multi-task attention network for automatic medical image segmentation and classification. *IEEE Transactions on Medical Imaging*, 43 (2), 674-685. <https://doi.org/10.1109/TMI.2023.3317088>
- [12] Zhu, M., Chen, Z., Yuan, Y. (2021). DSI-Net: Deep synergistic interaction network for joint classification and segmentation with endoscope images. *IEEE Transactions on Medical Imaging*, 40 (12), 3315-3325. <https://doi.org/10.1109/TMI.2021.3083586>
- [13] Celaya, A., Actor, J. A., Muthusivarajan, R., Gates, E., Chung, C., Schellingerhout, D., Riviere, B., Fuentes, D. (2023). PocketNet: A smaller neural network for medical image analysis. *IEEE Transactions on Medical Imaging*, 42 (4), 1172-1184. <https://doi.org/10.1109/TMI.2022.3224873>
- [14] Lee, M. C. H., Petersen, K., Pawlowski, N., Glocker, B., Schaap, M. (2019). TeTriS: Template transformer networks for image segmentation with shape priors. *IEEE Transactions on Medical Imaging*, 38 (11), 2596-2606. <https://doi.org/10.1109/TMI.2019.2905990>
- [15] Oktay, O., Ferrante, E., Kamnitsas, K., Heinrich, M., Bai, W., Caballero, J., Cook, S. A., de Marvao, A., Dawes, T., O'Regan, D. P., Kainz, B., Glocker, B., Rueckert, D. (2018). Anatomically constrained neural networks (ACNNs): Application to cardiac image enhancement and segmentation. *IEEE Transactions on Medical Imaging*, 37 (2), 384-395. <https://doi.org/10.1109/TMI.2017.2743464>
- [16] Kazemi, A., Keshtkar, A., Rashidi, S., Aslanabadi, N., Khodadad, B., Esmaeili, M. (2020). Automated segmentation of cardiac fats based on extraction of textural features from non-contrast CT images. In *2020 25th International Computer Conference, Computer Society of Iran (CSICC)*. IEEE. <https://doi.org/10.1109/CSICC49403.2020.9050072>
- [17] Zhang, Q., Zhou, J., Zhang, B., Jia, W., Wu, E. (2020). Automatic epicardial fat segmentation and quantification of CT scans using dual U-Nets with a morphological processing layer. *IEEE Access*, 8, 128032-128041. <https://doi.org/10.1109/ACCESS.2020.3008190>
- [18] Zlokolica, V., Velicki, L., Janev, M., Mitrinovic, D., Babin, D., Ralevic, N., Cemerlic-Adic, N., Obradovic, R., Galic, I. (2014). Epicardial fat registration by local adaptive morphology-thresholding based 2D segmentation. In *Proceedings ELMAR-2014*. IEEE. <https://doi.org/10.1109/ELMAR.2014.6923347>
- [19] Zhao, F., Hu, H., Chen, Y., Liang, J., He, X., Hou, Y. (2019). Accurate segmentation of heart volume in CTA with landmark-based registration and fully convolutional network. *IEEE Access*, 7, 57881-57893. <https://doi.org/10.1109/ACCESS.2019.2912467>

Received June 23, 2024
Accepted April 15, 2025