# A Dual-Stage Deep Learning Framework for Obesity Assessment with Ultrasound Image Segmentation and Severity Classification

**Rohini Pawar**
Department of Computer Engineering, Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune, India
rohini.pawar@bharatividyapeeth.edu (corresponding author)

**Rohini Jadhav**
Department of Information Technology, Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune, India
rbjadhav@bvucoep.edu.in

**Rohit Jadhav**
Department of ENT, Bharati Vidyapeeth (Deemed to be University) Medical College, Pune, India
dr.rohitjadhav@gmail.com

**Gagandeep Kaur**
Symbiosis Institute of Technology, Nagpur Campus, Symbiosis International (Deemed University), Pune, India
gagandeep.kaur@sitnagpur.siu.edu.in

**Rutuja Rajendra Patil**
Department of Computer Engineering, MIT Academy of Engineering, Pune, India
rutujapat@gmail.com

*Received: 26 October 2025 | Revised: 11 November 2025 and 18 November 2025 | Accepted: 19 November 2025*

 *DOI: https://doi.org/10.48084/etasr.15784*

**ABSTRACT**

**Obesity is a major global health challenge associated with metabolic dysfunction and multi-organ impairment. Early detection of fat accumulation is critical but depended on the resolution and operator dependency of conventional ultrasound techniques. This study proposes a two-stage deep learning framework for automated segmentation and severity classification of obesity using abdominal ultrasound images. The pipeline integrates four encoder–decoder architectures (ENet, SegNet, U-Net, and U-Net++) for precise fat-layer segmentation, followed by convolutional neural networks (LeNet-5, VGG19, ResNet50, and EfficientNetB0) for four-level obesity severity classification. A curated dataset of 550 radiologist-annotated ultrasound images was used for training and evaluation. The U-Net++ model achieved the highest segmentation performance with a Dice coefficient of 0.986 and an Intersection over Union of 0.973, while EfficientNetB0 recorded a superior classification accuracy of approximately 97% with stable validation performance. The results demonstrate that ultrasound-based deep learning pipelines can deliver reliable, non-invasive, and cost-effective solutions for early obesity detection, offering strong potential for integration into routine diagnostic practice.**

***Keywords-ultrasound imaging; obesity assessment; deep learning; U-Net++; EfficientNetB0; image segmentation; classification***

## I. INTRODUCTION

Obesity has reached epidemic proportions and is a major contributor to cardiometabolic disorders, hepatic dysfunction, and multi-organ morbidity across both developed and developing regions [1]. Beyond high body mass indices, ectopic and visceral fat accumulation cause progressive metabolic disturbances that precede the clinical symptoms [2]. These include hepatic steatosis, pancreatic fat deposition, myocardial remodeling, and renal changes, all of which elevate cardiometabolic risk. Hence, there is a need for diagnostic techniques that detect early focal fat infiltration before irreversible damage occurs. Recent studies show that subtle tissue and organ changes can be detected using advanced imaging and deep learning, for example, a 3D CT framework for visceral-fat obesity identified significant organ volume increases and attenuation decreases (pancreas $r = –0.701$, $p < 0.001$), supporting the potential of automated imaging biomarkers for early metabolic-risk detection, was proposed in [3]. Conventional diagnostic tools, including anthropometric measures, ultrasound scoring, CT, and MRI, have limitations in early-stage obesity assessment. Anthropometry lacks spatial accuracy, ultrasound is prone to operator dependence and speckle noise, and although CT and MRI provide high-fidelity quantification, they are impractical for routine screening due to cost and, in CT, radiation exposure [4]. These limitations hinder consistent detection of mild fat infiltration and reduce reproducibility across centers.

Artificial intelligence, particularly deep learning, provides objectivity, consistency, and fine-grained analysis in medical imaging. Encoder–decoder segmentation networks, such as U-Net and its variants (U-Net++, Attention U-Net, and U-Net 3+), effectively delineate low-contrast ultrasound structures using dense skip connections and multi-scale feature fusion [5, 6] Lightweight networks like ENet and SegNet enable real-time segmentation with improved boundary definition [7]. More recently, hybrid convolution–transformer architectures such as Swin-Transformer U-Net derivatives have improved segmentation by capturing long-range dependencies. For instance, the Swin-Net model proposed in [8] for breast ultrasound achieved Dice improvements of 1.4–1.8% over CNNs, demonstrating the potential of transformer-based designs for ultrasound fat analysis.

CNNs such as VGG, ResNet, and EfficientNet have shown strong performance in classifying hepatic and visceral fat infiltration. Transfer learning facilitates robust feature extraction from small datasets—VGG ensures stable convergence, ResNet offers superior generalization, and EfficientNet achieves optimal accuracy–efficiency balance [9, 10]. Despite these advances, most prior studies focus on binary classification with limited validation and small datasets [11]. The present study addresses these limitations through a dual-stage deep learning framework integrating segmentation (ENet, SegNet, U-Net, U-Net++) and classification (LeNet-5, VGG19, ResNet50, EfficientNetB0) to grade obesity severity from expert-annotated ultrasound images, thereby enhancing early detection, interpretability, and clinical applicability.

## II. METHODOLOGY

This study proposes a two-stage deep learning pipeline for ultrasound-based obesity assessment, combining (i) pixel-level segmentation of fat regions and (ii) four-class severity classification using segmented regions.

### A. Dataset Description

The dataset comprises approximately 550 expert-annotated ultrasound images in PNG format, each with corresponding binary masks delineating subcutaneous and visceral fat. Images without masks were excluded. The dataset (Table I) includes four classes: Normal, Mild, Moderate, and Severe [12].

TABLE I. DATASET SUMMARY

| Item | Value/note |
|---|---|
| **Total images** | ~550 (PNG) |
| **Classes** | 4 (Normal, Mild, Moderate, Severe) |
| **Annotations** | Expert-provided binary masks |
| **Access** | OSF DOI 10.17605/OSF.IO/C2YG8 (restricted) |

### B. Preprocessing

A standardized pipeline ensured consistent input quality, including format filtering, grayscale conversion, histogram-based contrast enhancement, thresholding (around 120), morphological filtering, normalization to [0,1], and resizing. Class-balancing augmentation was applied, and data were split 70:30 for segmentation and 80:10:10 for classification.

### C. Segmentation Models

Four encoder–decoder networks (ENet, SegNet, U-Net, and U-Net++) were trained using the Adam optimizer (learning rate = $1\times10^{-4}$) with learning rate scheduling and early stopping. ENet uses residual bottlenecks and dilated convolutions (1), SegNet performs pooled-index upsampling (2), U-Net applies symmetric skip connections (3), and U-Net++ introduces nested dense skips for refined boundary localization. The hybrid Binary Cross-Entropy and Dice Loss (4–5) improved segmentation accuracy. Evaluation metrics included Dice, IoU, precision, recall, F1-score, and accuracy.

$$X_{init} = \text{Concat}(\text{Conv}_{3\times3}^{s=2}(I), \text{MaxPool}_{2\times2}(I) \quad (1)$$

$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)] \quad (2)$$

$$\hat{Y} = \sigma(f_\theta(X)) \quad (3)$$

$$L(y,\hat{y}) = BCE(y,\hat{y}) + (1 - Dice(y,\hat{y})) \quad (4)$$

$$Dice(y,\hat{y}) = \frac{2\sum_i y_i\hat{y}_i + \epsilon}{\sum_i y_i + \sum_i \hat{y}_i + \epsilon} \quad (5)$$

TABLE II. SEGMENTATION TRAINING SETTINGS

| Model | Optimizer | LR | Loss | Epochs | Batch size |
|---|---|---|---|---|---|
| ENet | Adam | 1e-4 | BCE + Dice | up to 50 | 4 |
| SegNet | Adam | 1e-4 | BCE | up to 50 | 4 |
| U-Net | Adam | 1e-4 | BCE / Dice | up to 50 | 4 |
| U-Net++ | Adam | 1e-4 | BCE | up to 50 | 4 |

### *D. Classification Models*

Segmentation outputs from the best-performing model were classified into four severity levels using LeNet-5, VGG19, ResNet50, and EfficientNetB0. Models were trained with categorical cross-entropy using Adam or SGD optimizers (Table III). Performance was evaluated using accuracy, AUC, precision, and recall, with EfficientNetB0 achieving the highest accuracy.

TABLE III. CLASSIFICATION CONFIGURATIONS (REPRESENTATIVE)

| Model | Input | Optimizer | Typical LR |
|---|---|---|---|
| LeNet-5 | 128×128 | SGD | 1e-1 (as used) |
| VGG19 | 224×224 | Adam | 1.25e-5 |
| ResNet50 | 224×224 | Adam | 5e-5 |
| EfficientNetB0 | 224×224 | Adam | 2.5e-5 |

### *E. Experimental Workflow and Evaluation*

The overall experimental workflow followed a sequential process:

Ultrasound Data Acquisition → Preprocessing → Segmentation → Mask Generation → Data Construction → Classification → Evaluation.

Early stopping and learning rate scheduling were applied to all models to prevent overfitting and ensure generalization.

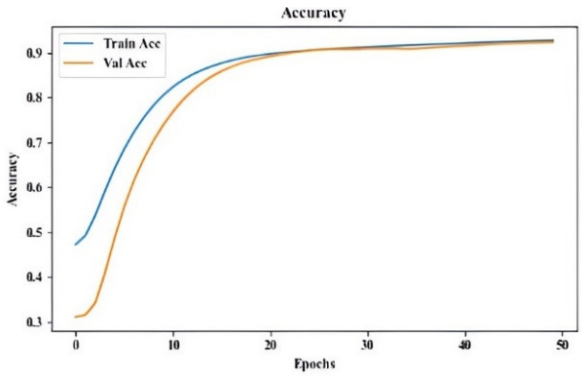
## III. RESULTS AND DISCUSSION

### *A. Segmentation Results*

Table IV presents the segmentation results, with U-Net++ achieving the highest Dice (0.9863) and IoU (0.9730). Accuracy–loss curves (Figures 1–2) show faster and more stable convergence, and the confusion matrices (Figures 3) reflect fewer misclassifications. The progressive improvement across models highlights the advantages of U-Net++'s nested skip connections and dense feature fusion, establishing it as the most effective segmentation backbone.

TABLE IV. SEGMENTATION PERFORMANCE

| Model | Dice | IoU | Precision | Recall | Accuracy |
|---|---|---|---|---|---|
| ENet | 0.9439 | 0.8938 | 0.9655 | 0.9233 | 0.9226 |
| SegNet | 0.9732 | 0.9478 | 0.9636 | 0.9829 | 0.9618 |
| U-Net | 0.9828 | 0.9661 | 0.9878 | 0.9778 | 0.9758 |
| U-Net++ | 0.9863 | 0.9730 | 0.9871 | 0.9856 | 0.9807 |



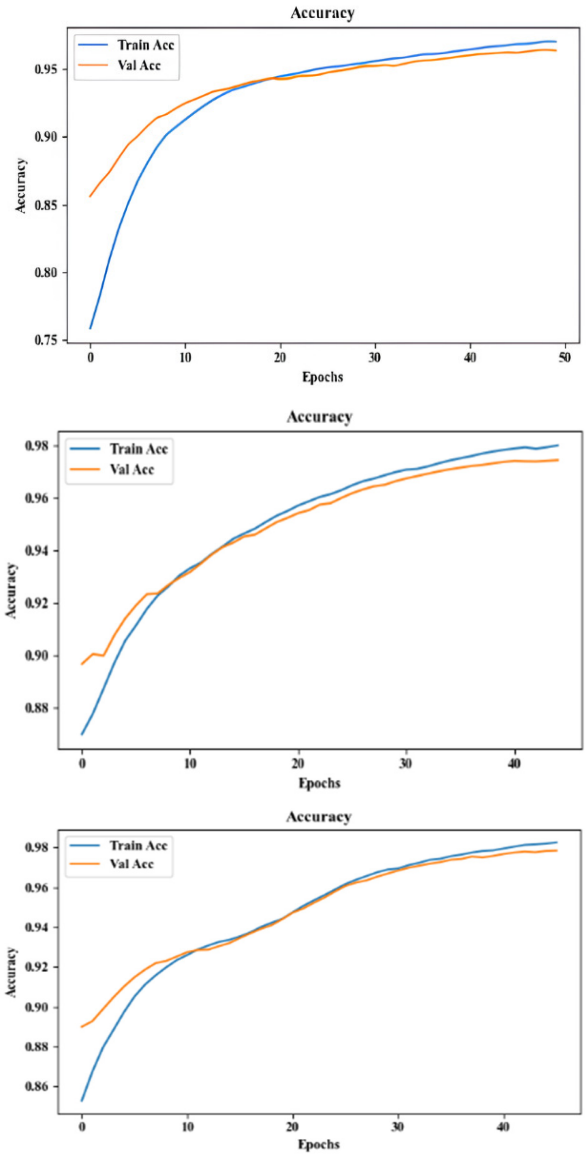
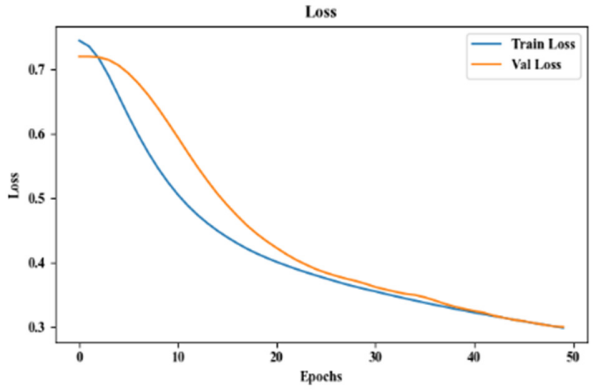




Fig. 1. Segmentation curves (accuracy) for ENet, SegNet, U-Net, and U-Net++.



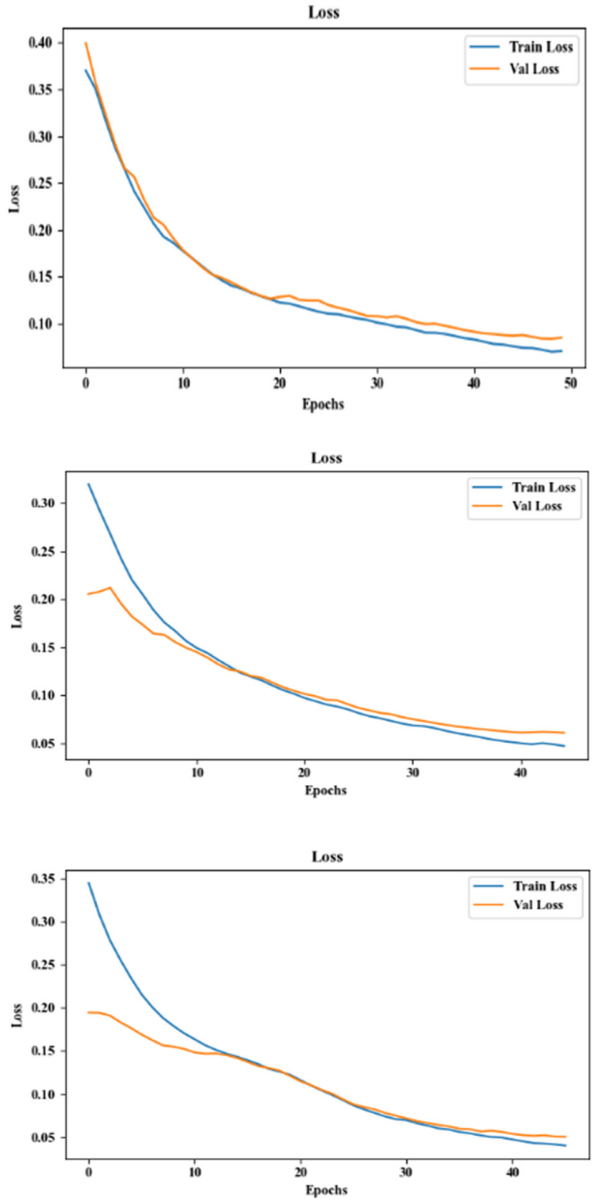
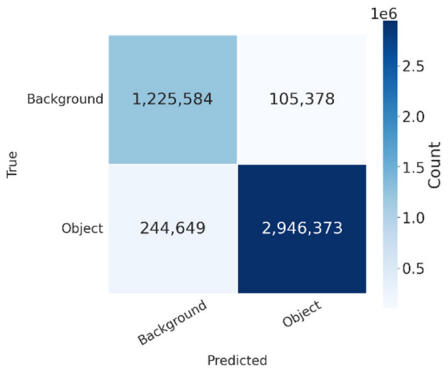



Fig. 2. Segmentation curves (loss) for ENet, SegNet, U-Net, and U-Net++.



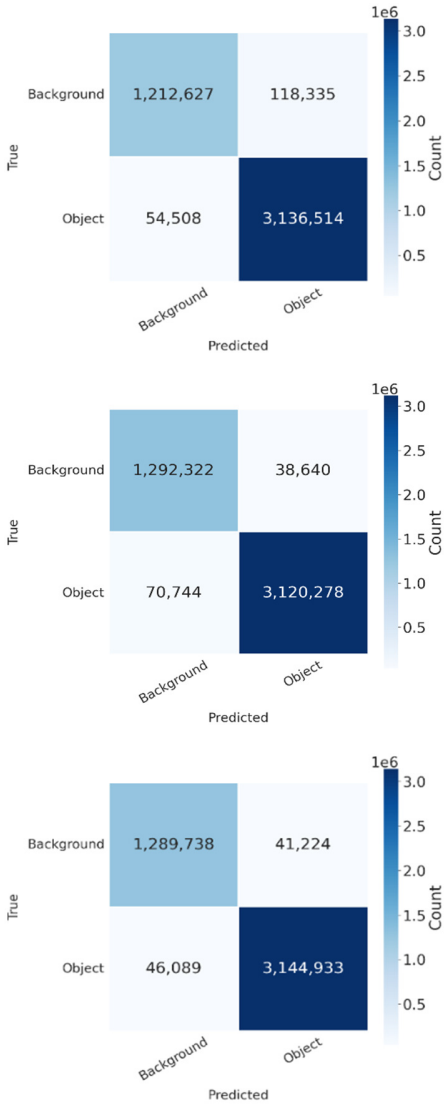




Fig. 3. Pixelwise confusion matrices for ENet, SegNet, U-Net, and U-Net++.

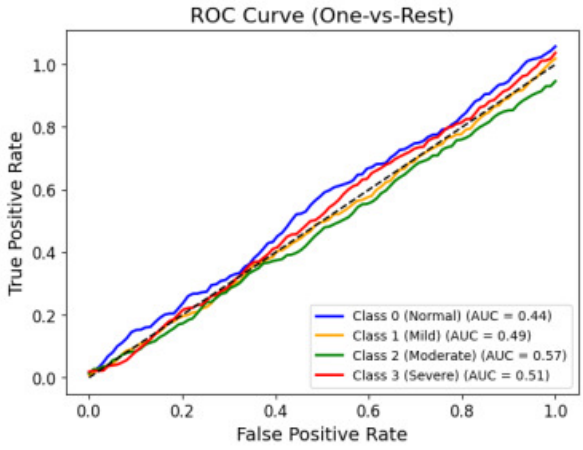
### *B. Classification Results*

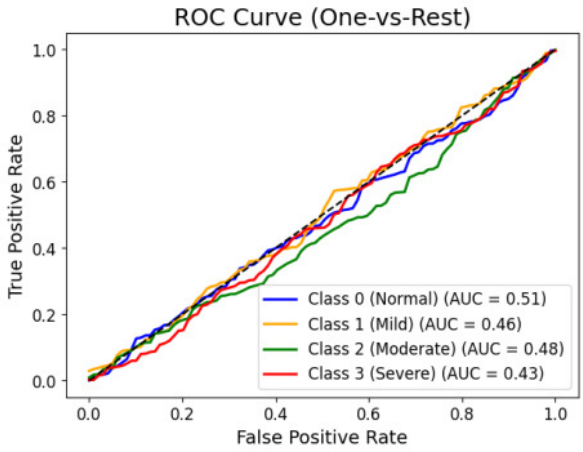
Table V summarizes the classification results. EfficientNetB0 achieved the highest accuracy and stability (training ≈ 97.3%, validation ≈ 96.5%), closely followed by VGG19. ResNet50 showed moderate performance, and LeNet-5 served as a low-capacity baseline (~30%). EfficientNetB0 and VGG19 accurately distinguished Normal and Severe classes, while Mild–Moderate overlap persisted (Figures 4–5), indicating reliable yet refinement-dependent multi-class performance.

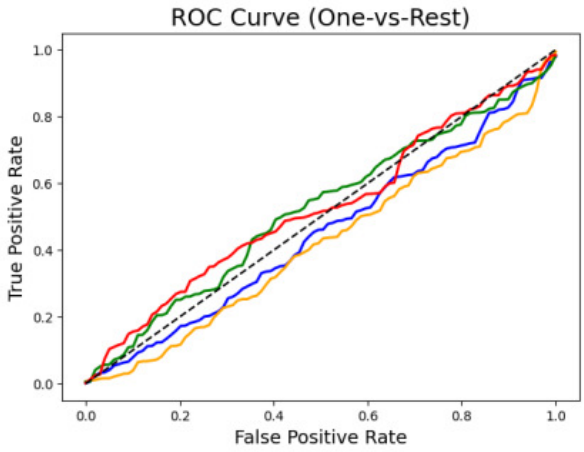
Across the related studies, only a few report directly comparable metrics, yet the available comparisons (Table VI) show that our dual-stage framework achieves superior segmentation precision and stronger multi-class classification performance, underscoring its effectiveness and clinical potential despite dataset and anatomical differences.

TABLE V. REPRESENTATIVE CLASSIFICATION SNAPSHOTS (SELECTED EPOCHS)

| Model | Representative training accuracy | Representative validation accuracy |
|---|---|---|
| LeNet-5 | ~27–29% | ~29% |
| VGG19 | ~97.7–98.2% | ~95–96% |
| ResNet50 | ~48–55% | ~71–77% |
| EfficientNetB0 | ~97.3% | ~96.5% |



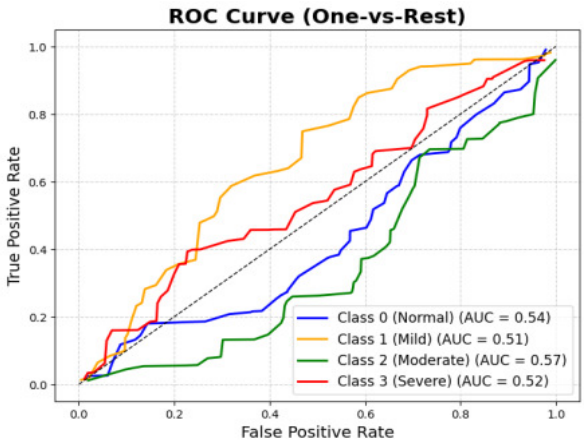
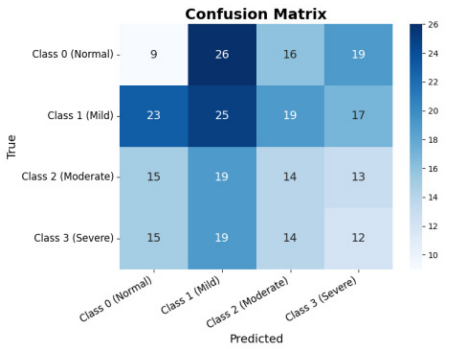
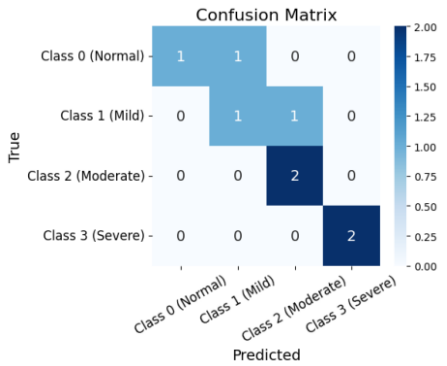
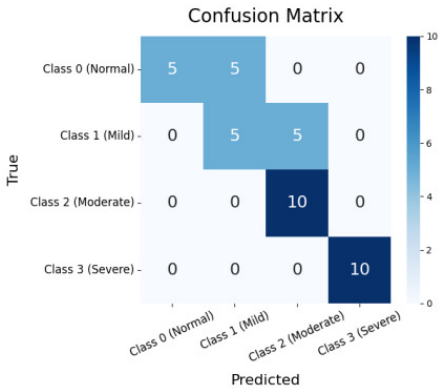




Fig. 4. One-vs-rest ROC/AUC curves for LeNet-5, VGG19, ResNet50, and EfficientNetB0 showing four obesity severity categories (Normal, Mild, Moderate, Severe).



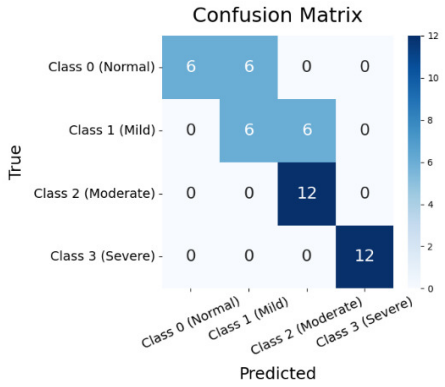



Fig. 5. Confusion matrices for LeNet-5, VGG19, ResNet50, and EfficientNetB0 illustrating inter-class relationships.

TABLE VI. PERFORMANCE COMPARISON WITH EXISTING STUDIES

| Study | Task / Modality | Dataset | Results |
|---|---|---|---|
| [13] | Ultrasound — gestational sac segmentation (different anatomy) | 500 ultrasound scans; 5-fold cross-validation | Dice = 0.978, IoU = 0.946 |
| [14] | Ultrasound — Subcutaneous Adipose Tissue (SAT) segmentation | 612 images (training) + 50 test images | Dice(SAT) = 0.9095 |
| [15] | CT — SAT segmentation (different modality) | Large public CT datasets | Dice(SAT) ≈ 0.9639 |
| [16] | Ultrasound — 4-class fatty-liver classification (normal/mild/moderate/severe) | Taiwan Biobank grouped images, subject-level bundles | 79.2% (4-class); 97.1% (binary) |
| Current study | Abdominal ultrasound — fat-layer segmentation + 4-class obesity-severity classification | ≈550 annotated images | Dice = 0.986, IoU = 0.973; Accuracy ≈ 96.5% |

## IV. CONCLUSION

This study developed a two-stage deep learning framework combining segmentation and classification for obesity severity assessment using ultrasound imaging. Encoder–decoder networks enabled precise fat-region delineation, while CNN classifiers accurately graded severity across 550 annotated images. U-Net++ achieved the best segmentation (Dice = 0.986), and EfficientNetB0 delivered the highest classification accuracy. The approach enhanced interpretability and diagnostic reliability, validating ultrasound as a non-invasive, cost-effective modality for AI-driven analysis. Despite the limited dataset size, the framework establishes a reproducible basis for multi-center validation and multimodal integration, supporting real-time deployment for clinical obesity assessment.

## REFERENCES

[1] "Obesity and overweight," *WHO*, Dec. 08, 2025. https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight.

[2] J. A. M. J. L. Janssen, "The Causal Role of Ectopic Fat Deposition in the Pathogenesis of Metabolic Syndrome," *International Journal of Molecular Sciences*, vol. 25, no. 24, Dec. 2024, https://doi.org/10.3390/ijms252413238.

[3] H. Kiyoyama *et al.*, "Association of visceral fat obesity with structural change in abdominal organs: fully automated three-dimensional volumetric computed tomography measurement using deep learning," *Abdominal Radiology*, vol. 50, no. 9, pp. 4395–4402, Sep. 2025, https://doi.org/10.1007/s00261-025-04834-x.

[4] T.-H. Chou *et al.*, "Deep learning for abdominal ultrasound: A computer-aided diagnostic system for the severity of fatty liver," *Journal of the Chinese Medical Association*, vol. 84, no. 9, Sep. 2021, Art. no. 842, https://doi.org/10.1097/JCMA.0000000000000585.

[5] S. M. Shaaban, M. Nawaz, Y. Said, and M. Barr, "An Efficient Breast Cancer Segmentation System based on Deep Learning Techniques," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12415–12422, Dec. 2023, https://doi.org/10.48084/etasr.6518.

[6] Y. K. Desai, "Diagnosis of Medical Images Using Convolutional Neural Networks," *Journal of Electrical Systems*, vol. 20, no. 6s, pp. 2371–2376, May 2024, https://doi.org/10.52783/jes.3220.

[7] P. Dong *et al.*, "An ultrasound image segmentation method for thyroid nodules based on dual-path attention mechanism-enhanced UNet++," *BMC Medical Imaging*, vol. 24, no. 1, Dec. 2024, Art. no. 341, https://doi.org/10.1186/s12880-024-01521-z.

[8] C. Zhu *et al.*, "Swin-Net: A Swin-Transformer-Based Network Combing with Multi-Scale Features for Segmentation of Breast Tumor Ultrasound Images," *Diagnostics*, vol. 14, no. 3, Jan. 2024, Art. no. 269, https://doi.org/10.3390/diagnostics14030269.

[9] A. Pati, S. R. Addula, A. Panigrahi, B. Sahu, D. S. K. Nayak, and M. Dash, "Artificial intelligence in improving disease diagnosis: A case study of cardiovascular disease prediction," in *Artificial Intelligence in Medicine and Healthcare*, A. Kumar, S. Rani, S. Rathee, N. Hemrajani, and M. Dahiya, Eds. CRC Press, 2025, https://doi.org/10.1201/9781003508595-2.

[10] T. Kim, D. H. Lee, E.-K. Park, and S. Choi, "Deep Learning Techniques for Fatty Liver Using Multi-View Ultrasound Images Scanned by Different Scanners: Development and Validation Study," *JMIR Medical Informatics*, vol. 9, no. 11, Nov. 2021, Art. no. e30066, https://doi.org/10.2196/30066.

[11] F. M. Alshagathrh, M. S. Househ, F. M. Alshagathrh, and M. S. Househ, "Artificial Intelligence for Detecting and Quantifying Fatty Liver in Ultrasound Images: A Systematic Review," *Bioengineering*, vol. 9, no. 12, Dec. 2022, Art. no. 748, https://doi.org/10.3390/bioengineering9120748.

[12] F. Alshagathrh *et al.*, "Large annotated ultrasound dataset of non-alcoholic fatty liver from Saudi hospitals for analysis and applications," *Data in Brief*, vol. 58, Feb. 2025, Art. no. 111266, https://doi.org/10.1016/j.dib.2024.111266.

[13] H. M. Danish, Z. Suhail, and F. Farooq, "Deep learning-based automation for segmentation and biometric measurement of the gestational sac in ultrasound images," *Frontiers in Pediatrics*, vol. 12, Dec. 2024, https://doi.org/10.3389/fped.2024.1453302.

[14] M. S. White, A. Horikawa-Strakovsky, K. P. Mayer, B. W. Noehren, and Y. Wen, "Open-Source AI for Vastus Lateralis and Adipose Tissue Segmentation to Assess Muscle Size and Quality," *Ultrasound in Medicine and Biology*, vol. 51, no. 12, pp. 2276–2280, Dec. 2025, https://doi.org/10.1016/j.ultrasmedbio.2025.08.008.

[15] M. Hayat, S. Aramvith, S. Bhattacharjee, and N. Ahmad, "Attention GhostUNet++: Enhanced Segmentation of Adipose Tissue and Liver in CT Images." arXiv, Apr. 14, 2025, https://doi.org/10.48550/arXiv.2504.11491.

[16] T.-J. Yen, C.-T. Yang, Y.-J. Lee, C. Chen, and H.-C. Yang, "Fatty liver classification via risk controlled neural networks trained on grouped ultrasound image data," *Scientific Reports*, vol. 14, no. 1, Mar. 2024, Art. no. 7345, https://doi.org/10.1038/s41598-024-57386-3.