International Journal of Basic and Applied Science 14 (4) (2026) 152-161

Published by: IOCSCIENCE

International Journal of Basic and Applied Science

Journal homepage: www.ijobas.pelnus.ac.id

Explainable Mitochondrial Image Segmentation and Morphological Quantification using Deep Learning Based Framework

Vandana Malik[1], A. J Singh[2]

[1,2] Computer Science Department, PH.D., Himachal Pradesh University, India

Article Info

Article history:

Received Dec 12, 2025
Revised Jan 21, 2026
Accepted Feb 24, 2026

Keywords:

Image Segmentation;
U-Net;
Grad-CAM;
Explainable AI;
Mitochondria.

ABSTRACT

Mitochondria is an essential cell organelle with varying shape and size. A slight change in mitochondrial morphology leads to neurodegenerative diseases. The advanced deep learning-based models like U-Net, Mark R-CNN, MitoNet, MitoStructSeg, MitoSkel perform accurate mitochondrial image analysis by performing image segmentation or morphological quantification but are devoid of the ability to interpret the results produced. This research work proposed a novel unified XM-DL framework (Explainable Mitochondrial Deep Learning Based Framework) capable of performing multiple tasks like image segmentation, morphological quantification, classification of mitochondria on the basis of their shape, and interpreting results by using explainable artificial intelligence (XAI) techniques as a single pipeline. The XM-DL framework is composed of U-Net architecture integrated with residual connections, skip connections, and attention gates for performing image segmentation, followed by a post processing module for morphological quantification and utilizing Gradient Class Activation Mapping (Grad-CAM) as explainable AI and form a unique pipeline. The XM-DL framework was trained on the MitoEM dataset and achieved a high F1 score of 0.9322 and IoU (intersection over union) of 0.8793 for image segmentation task. The XM-DL framework provides assistance to the medical service providers by improving the interpretability and understanding about the deep learning techniques.

Corresponding Author:

Vandana Malik,
Computer Science Department/PH.D.
Himachal Pradesh University
Summer Hill, Shimla 171005, India
Email: vandana@uithpu.ac.in

1. INTRODUCTION

Mitochondria is an essential cell organelle; it generates energy and is called the powerhouse of the cell [1]. It performs multiple functions like regulating metabolism, developing immunity, cell death (apoptosis), aging, etc. [2]. Mitochondria are dynamic in nature; they change their shape, structure, texture, and density due to changes in metabolic activities, mutation, stress, toxins, etc. [3]. Accurate and precise image segmentation of medical images and electron microscopy images plays an important role in the process of disease diagnosis and tracking cellular-level changes that occur during treatment [4]. The electron microscopy image of mitochondria is used to perform image segmentation to observe any changes in the morphology of the mitochondria. The morphological quantification provides

information about the varying shape and size of the mitochondria [5]. These morphological changes lead to neurodegenerative diseases like Alzheimer's, Parkinson's, bipolar disorder, schizophrenia, cancer, etc. [6]. Morphological quantification helps medical service providers to identify healthy and unhealthy mitochondria in the cells. The morphological quantification helps in keeping track of a patient's health conditions before and after the treatment [7]. Vital decisions regarding treatment planning and providing patients with optimum health benefits rely on the accuracy of medical data and how well it is interpreted by the doctor and healthcare service provider [8].

The Mask R-CNN model performs mitochondrial detection and image segmentation by using ResNet50 architecture integrated with a Feature Pyramid Network (FPN) and a region proposal network. This research work investigates the relationship between the shape of mitochondria and its functions. The presence of false negative results, generated by the model, needs post-processing [9]. The HIEV-Net model performs image segmentation of electron microscopy images by analyzing centerline details as shape indicators when annotated images are not available at a high scale and uses a hierarchical view ensemble convolution (HVEC) approach. The boundary of the cell organelles is not identified clearly, and it hampers the segmentation results [10]. The EM-Net model is a deep convolutional neural network used for image segmentation of mitochondria, and it is based upon U-Net architecture. With limited data as well, the EM-Net model was able to perform well. The EM-Net tried to overcome the issue of low-contrast images and limited ground truth data availability [11]. The Mask R-CNN, HIEV-Net, and EM-Net models only perform image segmentation tasks but lack the ability to provide explanations about the segmented results.

The fully convolutional neural network architecture performs image segmentation and quantitative measurements of mitochondria and uses synthetic images, which are unable to represent mitochondria in a realistic manner. It has outperformed the adaptive active mask (AAM) technique used earlier, and it involved deformation area similarity and deformation distance similarity metrics to determine sensitivity [12]. The MitoNet model performs instance image segmentation and morphological quantification of mitochondria to inculcate an understanding about the structures and functions of mitochondrial cell organelles by utilizing the Empanada Python library. The MitoNet struggles with false negatives as it fails to detect smaller mitochondria [13]. The PHILOW Python-based human-in-the-loop workflow is designed for segmentation and analysis of mitochondria cristae. HITL (human in the loop) helps in correcting and improving the accuracy of image segmentation manually. The dependency on humans has increased as they are involved for proofreading to enhance the accuracy of the platform [14]. The MoDL is a deep learning algorithm used for performing image segmentation of live cells and predicting the functions of mitochondria according to their varying shapes. During cancer treatment, the shape of mitochondria varies, and the changes are captured by MoDL and analyzed for mitochondrial morphology. The ensemble learning increases the computational cost, but it also leads to better prediction accuracy, and it justifies the trade-off [15]. The MitoStructSeg model is used to perform quantitative analysis and segmentation of mitochondria by incorporating Adaptive Multidomain Mitochondrial Segmentation (AMM-Seg Model). The MitoStructSeg is a user-friendly tool used for analyzing the health of mitochondria and cristae structure. Changes in inner and outer cristae surface area indicate damage caused by diseases and harm healthy mitochondria. The MitoStructSeg model faces challenges like data variability and the inability to manage data from different imaging technologies; it needs to process unlabeled data [16]. MitoSkel is an artificial intelligence tool developed for automatic image segmentation and mitochondrial quantification. It incorporates U-Net with a Gabor Filter Layer and Thresholding Attention Mechanism (TAM). Mitochondrial quantitative analysis is performed by using skeletonization, and parameters like length, branching, and connectivity are extracted. The precise delineation of overlapping mitochondria is a challenging task, especially in low-resolution images [5].

The existing models are categorized as models performing image segmentation task only, like Mask R-CNN, HIEV-Net, and EM-Net models. Other models like MitoNet, MitoStructSeg, MitoSkel, and MoDL are performing morphological quantification along with the image segmentation task.

These models provide crucial useful information about the mitochondrial morphology but without giving any explanation about how the results were generated.

After the literature review of image segmentation and morphological quantification of mitochondria, it was observed that there is dependency on human experts to segment data more accurately and precisely, but it consumes lots of time and effort. There are models that are unable to cope with the dynamic and varying nature of mitochondria. The small cellular overlapping structures hamper the mitochondrial image segmentation. Some existing models are performing only the image segmentation task, and some are performing the image segmentation task along with morphological quantification, but no one performs these tasks in a unified manner, and neither incorporates explainable AI into the model for justifying the results produced. The medical professionals do not understand the working of these deep learning models and find it difficult to interpret and trust the results produced and decisions made [17].

The main contribution of this research work is to make a deep learning-based framework capable of performing multiple tasks as a single process. It performs image segmentation, morphological quantification, classification of mitochondria on the basis of their shape, counting the number of mitochondria present in an input image, and incorporating an explainable AI component into the framework for clearly showing the region involved for generating the results and creating a unified pipeline for performing multiple tasks.

The complex deep learning models work with a huge number of parameters; each iteration calculates some intermediate results, and only the final results are showcased, but without any justification or explanation. The medical professionals hesitate from utilizing these deep learning models while giving treatment to the patients because of their ambiguous nature. The Explainable AI (XAI) bridges this gap by explaining the complex deep learning models and specifying how the algorithms extract results from the given data. It helps the medical professionals understand these deep learning technologies and how they work. The XAI is capable of building trust, transparency, and interpretability among the medical professionals [18]. Some of the explainable AI methods used are Grad-CAM (Gradient-weighted Class Activation Mapping), Local Interpretable Model-agnostic Explanation (LIME), and SHapley Additive explanation (SHAP). In this research work the Grad-CAM is utilized for interpreting the results through visualization because it performs well with CNN-based models and indicates where the model paid its attention while performing the task [19]. The LIME or SHAP methods provide image level or feature level explanations, whereas Grad-CAM works with convolutional feature maps, and it is suitable for dense segmentation tasks.

2. RESEARCH METHOD

The grayscale electron microscopy (EM) images of mitochondria from the publicly available MitoEM dataset were used for training and testing the proposed framework. The MitoEM dataset is composed of MitoEM-H (human) and MitoEM-R (rat) volumes with size (1000, 4096, 4096) voxels for the (x, y, z) axis. The MitoEM-H dataset images are utilized in this research work. There are 400, 100, and 500 images in their train, validation, and test folders respectively, in the .png extension, and their corresponding labels in the .tif format are also available separately [20]. The volumetric electron microscopy data is present in the MitoEM dataset, but 2D slices were used because of the availability of 2D annotated data. This work does not handle the interslice dependencies and three-dimensional morphology of the mitochondria, but in the future, this can be explored.

This research work started with the preprocessing steps; the images were converted into .jpg format and resized into 256 x 256 frame size. The images were saved into their respective folders. The original images were complex and big in size; after resizing them, 102400 frames were created for training. Similarly, their masks were also converted into .jpg format from .tif format and resized into 256 x 256 dimension. A similar process is followed by the testing images as well, and finally 128000 test image frames are created along with their 128000 true mask images. As these images are huge in number, and due to the resource-constraint environment and limited computational power, only a limited number of images were used. For the training process, 5000 image frames along with their true

masks were used, out of which 4000 images were used to train the model and 1000 images were used for validation. For testing, 2000 images were taken along with their true masks for evaluating the performance of the model.

The U-Net model is specially designed for medical image analysis, based upon CNN architecture [21]. The encoder and decoder are two building blocks of the U-Net model. The encoder part captures features and context from the images through convolutional layers and max pooling and performs downsampling. There is a bottleneck layer containing the most abstract information about the image, and it connects the encoder to the decoder. The decoders perform the reconstruction of the image step by step by upsampling the features passed from the lower layers to build the original image. The skip connections link the encoder to the decoder to directly pass and preserve the fine details about the image at each step. The feature maps passed from the lower convolutional layers are combined with the details provided by the skip connection for precise localization and produce accurate results [22] . To the basic U-Net model, the residual connections are incorporated because electron microscopy cell images are very complex, with varying shapes and textures of mitochondria. The residual connections help the network to learn crucial features, remember important spatial details, and improve gradient flow [23]. The skip connection is a vital part of the U-Net model; attention gates are utilized to filter the features obtained by skip connection. The attention gates emphasize the most relevant regions of the image and suppress the irrelevant parts. The features of high importance are sent forward by the attention gates to precisely improve the image segmentation task [24].

In this research work the mitochondrial image segmentation is followed by a post-processing step performing morphological quantification of the mitochondrial structures. The deep learning models provide no explanation about the results they produce or how they achieved them. Therefore, the doctors, researchers, and medical service providers lack trust and understanding about these models, as their results may be influenced or biased due to data imbalance [25]. The Grad-CAM is utilized in the post processing steps after image segmentation to analyze the results generated by the proposed models. It interprets the results to develop understanding and build trust among the users by emphasizing the region, playing major role in producing the results and depicting them through heatmaps.

The implementation of this research work is performed in an environment with the following hardware specification: processor of Intel(R) Core (TM) Ultra 9 185H, 2300 MHz, 16 Core(s), 22 Logical Processor(s), Installed Physical Memory (RAM): 32.0 GB, Operating System: Microsoft Windows 11. Software like Anaconda with Jupyter Notebook and Python is used for compiling and executing the code. The deep learning frameworks like PyTorch and TensorFlow are used. Libraries like NumPy, pandas, matplotlib, skimage, etc., are used for performing various tasks and functions during implementation.

The proposed XM-DL framework (Explainable Mitochondrial Deep Learning Based Framework) is a unique and unified pipeline performing mitochondrial image segmentation followed by morphological quantification and explainable AI component for helping understand the results generated by the framework. It is based upon U-Net architecture incorporated with an attention mechanism and residual connections for performing accurate segmentation, and it is six layers deep. The residual connections are utilized for reducing vanishing gradients as the layers get deeper in the network. It also enables the model to segment edges of mitochondria accurately. The encoder path of the architecture has convolutional layers for feature extraction followed by batch normalization and max pooling for downsampling. The encoder and decoder are connected to each other with the help of a bottleneck layer containing essential feature details of the image. The important spatial features are transferred from the encoder to the decoder side through skip connections, which are enhanced by the attention mechanism. It improves the performance of the model by reducing false negative results and efficiently detecting the low-contrast mitochondria. The decoder part performs upsampling by using transposed convolution. Then the results from residual refinements are concatenated with features passed through attention and skip connections. The pixel-wise class

probabilities are determined by the Softmax activation function. This U-Net architecture is six layers deep because it extracts multiple features from the mitochondrial images accurately. The shallow architectures generate false negative outputs, whereas deeper architectures face overfitting problems and produce false positive results. Therefore, the six layer deep architecture balances the contextual abstraction and spatial resolution to produce accurate results during mitochondrial image analysis. The training of the proposed framework was set to 50 epochs, but due to the utilization of early stopping with patience=5, the model got trained up to 22 epochs because validation loss did not improve after that. It avoids the model form getting overfitted, and it preserves the weights of the best performance. The Adam optimizer with learning rate of 0.001 is used for stable convergence in image segmentation tasks. The Dice Loss is used to handle class imbalance in medical images by emphasizing boundary overlap. The performance of the model is evaluated on the parameters like F1 score and IoU. The post-processing steps, like mitochondrial morphological quantification and explainable AI, are applied after the framework performs image segmentation. The morphology of the mitochondria is quantified in terms of area, perimeter, major axis, minor axis, eccentricity, solidity, and aspect ratio [5]. The mitochondria are also classified on the basis of their shapes, like round, elongated, tubular, fragmented, and irregular. These measurements help in tracking the mitochondrial health during treatment. The Grad-CAM forms the explainable AI module in the post-processing phase of the XM-DL framework. The Grad-CAM generates heatmaps according to specific class and indicates the most significant part of the image that influences the segmentation results. The Figure 1. is a detailed diagram of the XM-DL framework indicating the workflow starting from image pre-processing to image segmentation followed by morphological quantification and explainable AI using Grad-CAM and displaying the unified visualization results along with morphological quantification details in the table.

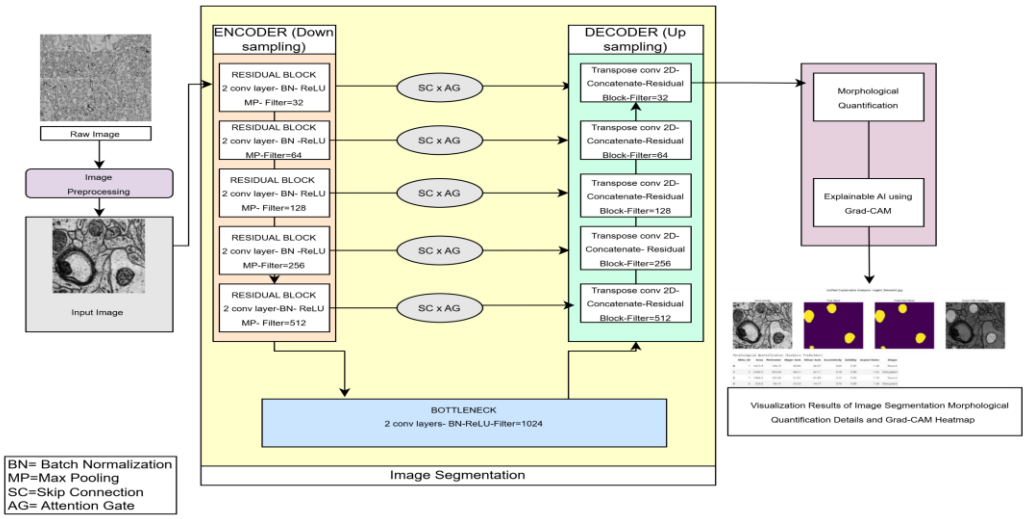
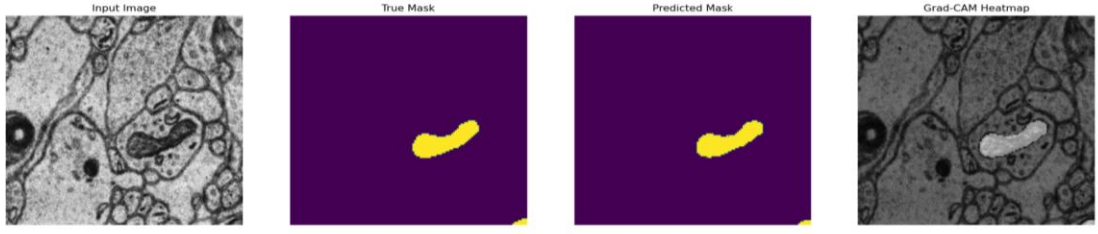


Figure 1. The XM-DL framework indicating the workflow and the final results

3. RESULTS AND DISCUSSIONS

The XM-DL framework is trained on the MitoEM dataset for 22 epochs and obtains a high training accuracy of 0.9924, a loss of 0.0077, a precision of 0.9924, and a recall of 0.9924. The validation accuracy is 0.989, the validation loss is 0.0107, the validation precision is 0.9894, and the validation recall is 0.9894. These values indicate that the model is well trained, is learning effectively, and is fitting well on training data and performing well on validation data. The performance of the XM-DL framework is evaluated on parameters like F1 score = 0.9322 and IoU=0.8793 for the image segmentation task. These parameters evaluations indicate that accurate image segmentation is performed, and low false negative and false positive results are obtained.

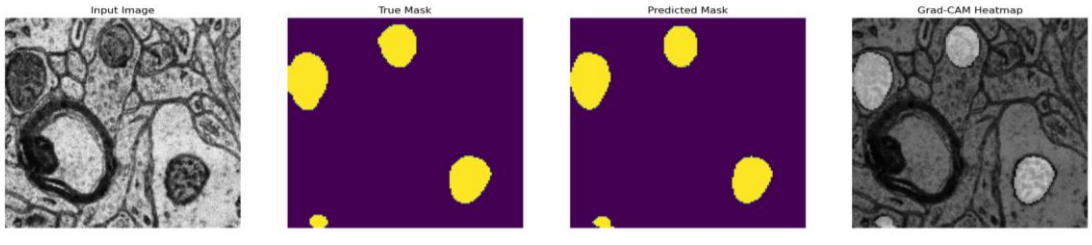
The post-processing modules of the XM-DL framework are capable of performing morphological quantification on the segmented images. The input image and its true mask are given to the framework. The predicted mask is generated, indicating the mitochondria as yellow pixels in the image and the background in black pixels. The morphological features like area, perimeter, major axis, minor axis, eccentricity, solidity, aspect ratio, and shape are quantified in CSV format. The morphological features are evaluated by utilizing regionprops from the skimage library; it is a standard bioimage analysis tool utilizing mathematical operations and deterministic functions for making accurate measurements. The numerical description of the mitochondrial geometry is obtained and summarizes the segmentation results into a biologically interpretable form. These measurements are used in the rule-based mitochondrial shape classification by comparing the obtained measurements to the threshold values and classifying the mitochondria as round, elongated, fragmented, tubular, or irregular in shape. While monitoring the patient's health and observing the effect of drugs and treatment on the mitochondria can be tracked by taking into account these morphological quantification values. The morphological quantification is dependent upon the accuracy of the images segmentation task because the impact of false negative and false positive results during image segmentation leads to unreliable morphological quantification. The explainable AI module utilizes Grad-CAM in the XM-DL framework. After the image segmentation is done, it does not influence the training process nor modify the loss function; it performs the qualitative inspection of the results produced. The Grad-CAM generates class specific visualization to highlight the region of interest being focused by the attention mechanism of the framework. These visualizations are important in medical image segmentation tasks because incorrect attention and focus may lead to false positives and unreliable outcomes. The Grad-CAM helps the researchers to access the spatial sensitivity and know whether the framework is working as per the domain expectations or not. The Grad-CAM takes the feature maps from the last convolutional layer by computing the gradients and generates a coarse localization map. The spatial region that contributed to make the prediction is highlighted through the heatmaps [26]. Here the Grad-CAM heatmaps highlight the region majorly contributing to producing mitochondrial image segmentation. If the predicted mask coincides with the heatmaps produced by the Grad-CAM, it means that the model is well trained and the model is focusing on the relevant parts of the image. The Grad-CAM brings visual interpretability and develops understanding about image segmentation results. The Figure 2. Figure 3. and Figure 4. are the results produced by the XM-DL framework, clearly indicating a single output showcasing all the results at once in a unified manner for image segmentation, morphological quantification details, and Grad-CAM heatmap.



Morphological Quantification (Explains Prediction)

	Mito_ID	Area	Perimeter	Major Axis	Minor Axis	Eccentricity	Solidity	Aspect Ratio	Shape
0	1	1617.0	222.36	86.89	27.68	0.95	0.81	3.14	Tubular
1	2	69.0	34.45	14.82	6.29	0.91	0.95	2.36	Elongated

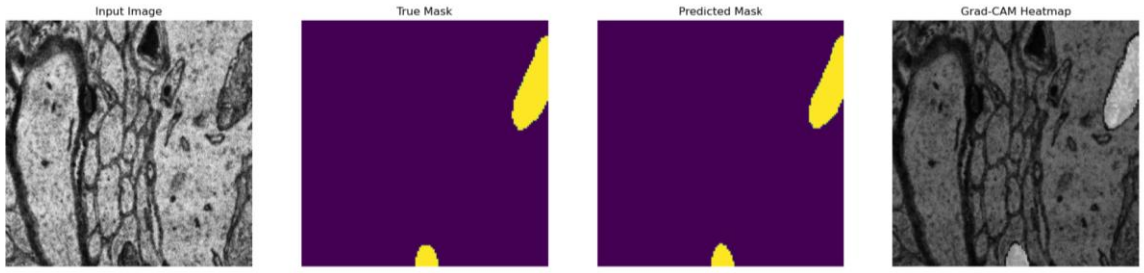
Figure 2. Visualization of the XM-DL framework result showing the input image true mask predicted mask and Grad-CAM Heatmap morphological quantification details and shape classification of mitochondria as tubular and elongated



Morphological Quantification (Explains Prediction)

	Mito_ID	Area	Perimeter	Major Axis	Minor Axis	Eccentricity	Solidity	Aspect Ratio	Shape
0	1	1473.0	158.73	50.85	36.97	0.69	0.97	1.38	Round
1	2	2306.0	205.84	69.21	42.71	0.79	0.96	1.62	Elongated
2	3	1868.0	182.60	57.81	41.49	0.70	0.95	1.39	Round
3	4	224.0	65.11	21.29	14.17	0.75	0.89	1.50	Elongated

Figure 3. Visualization of the XM-DL framework result showing the input image true mask predicted mask and Grad-CAM Heatmap morphological quantification details and shape classification of mitochondria as round and elongated



Morphological Quantification (Explains Prediction)

	Mito_ID	Area	Perimeter	Major Axis	Minor Axis	Eccentricity	Solidity	Aspect Ratio	Shape
0	1	1991.0	243.08	96.38	28.33	0.96	0.93	3.4	Tubular
1	2	423.0	86.14	25.29	23.02	0.41	0.93	1.1	Round

Figure 4. Visualization of the XM-DL framework showing mitochondria in the predicted mask and Grad-CAM Heatmap overlapping with the True mask image indicating accurate segmentation along with morphological quantification

In Table 1. the multitasking ability of the existing models and the proposed framework are compared. The tasks like Image Segmentation (IS), Morphological Quantification (MQ), Explainable Artificial Intelligence (XAI), and Classification (C) are performed by these models. It is difficult to directly compare the performance of the existing models because different types of mitochondrial images, like 2D and 3D, are used; different datasets are used, different tasks are performed; and different parameters are used to evaluate the performance of the models.

Table 1. Comparing the existing models on the basis of the tasks performed and performance metrics

Ref	Model / Year	IS	MQ	C	XAI	Performance Metrics
[9]	Improved Mask R-CNN 2018	Yes	No	No	No	Detection (IoU≥0.7) FIB-SEM- Prec= 0.872/ 0.930, Rec= 0.894/ 0.842 ATUM-SEM-Prec= 0.812/ 0.922, Rec= 0.952/ 0.925 Segmentation (IoU) FIB-SEM= 0.849 ATUM-SEM=0.864

Ref	Model / Year	IS	MQ	C	XAI	Performance Metrics
[12]	FCN 2018	Yes	Yes	No	No	FCN AS=0.0112, U-Net= 0.0236, ABD: FCN= 0.0466, U-Net=0.1023, Sensitivity: DAS-U-Net=0.5581, DDS-UNet=0.8788
[13]	MitoNet 2023	Yes	Yes	No	No	F_{1}@75 =0.88 (Lucchi++) Semantic IoU=0.79-0.90 and Panoptic Quality (PQ)=0.57-0.83
[16]	MitoStructSeg 2024	Yes	Yes	No	No	High F_{1}/IoU score but exact values are not specified. F_{1} score is ranging from 0.3 to 0.5 higher than the other methods.
[15]	MoDL 2025	Yes	Yes	No	No	Dice=0.92 mIoU=0.84 PA=0.95
Our	XM-DL 2025	Yes	Yes	Yes	Yes	F_{1}=0.9322, IoU=0.8793

The XM-DL framework is superior to the existing models because it performs multiple tasks like image segmentation, morphological quantification, shape classification, count the number of mitochondria in the input image, and heatmap visualization using Grad-CAM to explain the most significant region of the image responsible for generating the results in a single unified process. It also achieved high F_{1} score and IoU for the image segmentation task. There are some limitations of the XM-DL framework, like the smaller dataset that is used because of the constrained experimental setup and the ablation study that is not conducted. Only 2D images of mitochondria are used for the implementation purpose, and during the conversion of 3D images to 2D, the inter-slice contextual information is lost. The Grad-CAM provides a very generalized and coarse visual explanation by highlighting the region; it fails to give pixel-level information.

4. CONCLUSION

The XM-DL is an efficient framework performing detailed analysis of a vital cell organelle called mitochondria. The XM-DL framework was able to combine multiple tasks like image segmentation, morphological quantification, and explainable AI in a single process. The varying shapes of mitochondria were accurately segmented, and their geometric dimensions were recorded for helping the medical professionals while making diagnoses. The Grad-CAM efficiently highlights the area contributing to making major predictions. It helps medical professionals to understand the working of the framework. The complete overlap between the true mask, predicted results, and Grad-CAM heatmaps indicates that the framework has performed accurately. The explainability and interpretability of deep learning-based results increases its chances of being accepted and utilized by the medical professionals for performing clinical tasks. Currently the XM-DL framework was trained on 2D images, but in the future, a large 3D image dataset can be utilized for complete analysis of the mitochondrial morphology. Other explainable AI techniques can also be explored for obtaining more fine-grained pixel-level explanations. Other cell organelles apart from mitochondria, like the nucleus, lysosomes, Golgi apparatus, etc., can also be analyzed through the XM-DL framework. The XM-DL framework can be evaluated on multiple datasets and can be improved by using lightweight techniques. The deep learning technologies are now becoming an integral part of healthcare services. Therefore, it is very important to build trustworthy systems and incorporate transparency in deep learning frameworks and models. The explainable AI ensures that medical professionals understand the internal operations of the framework, so they can incorporate these techniques into their routine practices, which helps in transforming the healthcare services.

ACKNOWLEDGEMENTS

I would like to thank my guide, Prof. A.J Singh for providing supervision and guidance from time to time. I would also like to thank my friends and family for believing in me and providing unwavering support.

DECLARATIONS

AI USAGE STATEMENT

I declare that AI-assisted technology does not influence the research work. No part of the conceptual framework, data interpretation, or conclusions was generated by AI. The authors have thoroughly reviewed and verified all content, and accept full responsibility for the work presented.

AUTHOR CONTRIBUTION

The research work presented in this manuscript was conducted by Vandana Malik, as a PH.D. research scholar and Professor A. J Singh provided continuous guidance and supervision throughout the project.

FUNDING STATEMENT

The research presented in this manuscript is not funded by any agency, organization, or institute.

CONFLICTING INTERESTS

The authors of this manuscript i.e. Vandana Malik and Prof. A.J Singh, have no conflict of interest nor there is any financial or non-financial interest in the submitted work.

REFERENCES

[1] M. Changaei *et al.*, "From powerhouse to modulator: regulating immune system responses through intracellular mitochondrial transfer," *Cell Commun Signal*, vol. 23, no. 1, p. 232, May 2025, doi: 10.1186/s12964-025-02237-5.

[2] "Deep Analysis of Mitochondria and Cell Health Using Machine Learning | Scientific Reports." Accessed: Dec. 17, 2024. [Online]. Available: https://www.nature.com/articles/s41598-018-34455-y

[3] B. C. Jenkins *et al.*, "Mitochondria in disease: changes in shapes and dynamics," *Trends in Biochemical Sciences*, vol. 49, no. 4, pp. 346–360, Apr. 2024, doi: 10.1016/j.tibs.2024.01.011.

[4] B. Abhisheka, S. K. Biswas, B. Purkayastha, D. Das, and A. Escargueil, "Recent trend in medical imaging modalities and their applications in disease diagnosis: a review," *Multimed Tools Appl*, vol. 83, no. 14, pp. 43035–43070, Oct. 2023, doi: 10.1007/s11042-023-17326-1.

[5] S. Zaghbani, R. K. Pranti, L. Faber, and A. J. Garcia-Saez, "MitoSkel: AI tool for semantic segmentation and quantification of mitochondria from light microscopy images," *Biomedical Signal Processing and Control*, vol. 106, p. 107762, Aug. 2025, doi: 10.1016/j.bspc.2025.107762.

[6] A. Chaddad, Y. Hu, Y. Wu, B. Wen, and R. Kateb, "Generalizable and explainable deep learning for medical image computing: An overview," *Current Opinion in Biomedical Engineering*, vol. 33, p. 100567, Mar. 2025, doi: 10.1016/j.cobme.2024.100567.

[7] R. Gipiškis, C.-W. Tsai, and O. Kurasova, "Explainable AI (XAI) in image segmentation in medicine, industry, and beyond: A survey," *ICT Express*, vol. 10, no. 6, pp. 1331–1354, Dec. 2024, doi: 10.1016/j.icte.2024.09.008.

[8] D. Cheng, Z. Zhou, H. Li, J. Zhang, and Y. Yang, "A morphological difference and statistically sparse transformer-based deep neural network for medical image segmentation," *Applied Soft Computing*, vol. 174, p. 113052, Apr. 2025, doi: 10.1016/j.asoc.2025.113052.

[9] J. Liu, W. Li, C. Xiao, B. Hong, Q. Xie, and H. Han, "Automatic Detection and Segmentation of Mitochondria from SEM Images using Deep Neural Network," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Jul. 2018, pp. 628–631. doi: 10.1109/EMBC.2018.8512393.

[10] Z. Yuan, X. Ma, J. Yi, Z. Luo, and J. Peng, "HIVE-Net: Centerline-aware hierarchical view-ensemble convolutional network for mitochondria segmentation in EM images," *Computer Methods and Programs in Biomedicine*, vol. 200, p. 105925, Mar. 2021, doi: 10.1016/j.cmpb.2020.105925.

[11] A. Khadangi, T. Boudier, and V. Rajagopal, "EM-net: Deep learning for electron microscopy image segmentation," in *2020 25th International Conference on Pattern Recognition (ICPR)*, Jan. 2021, pp. 31–38. doi: 10.1109/ICPR48806.2021.9413098.

[12] X. Chai, Q. Ba, and G. Yang, "Characterizing robustness and sensitivity of convolutional neural networks for quantitative analysis of mitochondrial morphology," *Quantitative Biology*, vol. 6, no. 4, pp. 344–358, 2018, doi: 10.1007/s40484-018-0156-3.

[13] R. Conrad and K. Narayan, "Instance segmentation of mitochondria in electron microscopy images with a generalist deep learning model trained on a diverse dataset," *Cell Systems*, vol. 14, no. 1, pp. 58-71.e5, Jan. 2023, doi: 10.1016/j.cels.2022.12.006.

[14] S. Suga, K. Nakamura, Y. Nakanishi, B. M. Humbel, H. Kawai, and Y. Hirabayashi, "An interactive deep learning-based approach reveals mitochondrial cristae topologies," *PLOS Biology*, vol. 21, no. 8, p. e3002246, Aug. 2023, doi: 10.1371/journal.pbio.3002246.

[15] Y. Ding *et al.*, "Mitochondrial segmentation and function prediction in live-cell images with deep learning," *Nat Commun*, vol. 16, no. 1, p. 743, Jan. 2025, doi: 10.1038/s41467-025-55825-x.

[16] X. Wang *et al.*, "MitoStructSeg: A Comprehensive Platform for Mitochondrial Structure Segmentation and Analysis," Jul. 02, 2024, *bioRxiv*. doi: 10.1101/2024.06.28.601295.

[17] V. Tucci, J. Saary, and T. E. Doyle, "Factors influencing trust in medical artificial intelligence for healthcare professionals: a narrative review," *Journal of Medical Artificial Intelligence*, vol. 5, no. 0, Mar. 2022, doi: 10.21037/jmai-21-25.

[18] A. Chaddad, Y. Hu, Y. Wu, B. Wen, and R. Kateb, "Generalizable and Explainable Deep Learning for Medical Image Computing: An Overview," *Current Opinion in Biomedical Engineering*, vol. 33, p. 100567, Mar. 2025, doi: 10.1016/j.cobme.2024.100567.

[19] J. Narkhede, "Comparative Evaluation of Post-Hoc Explainability Methods in AI: LIME, SHAP, and Grad-CAM," in *2024 4th International Conference on Sustainable Expert Systems (ICSES)*, Oct. 2024, pp. 826–830. doi: 10.1109/ICSES63445.2024.10762963.

[20] "MitoEM dataset: large-scale 3d mitochondria instance segmentation (Paper) — BiaPy documentation." Accessed: Nov. 17, 2025. [Online]. Available: https://biapy.readthedocs.io/en/latest/tutorials/instance_seg/mitoem.html

[21] R. Wang, T. Lei, R. Cui, B. Zhang, H. Meng, and A. K. Nandi, "Medical image segmentation using deep learning: A survey," *IET Image Processing*, vol. 16, no. 5, pp. 1243–1267, Apr. 2022, doi: 10.1049/ipr2.12419.

[22] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., Cham: Springer International Publishing, 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.

[23] T. Hussain and H. Shouno, "MAGRes-UNet: Improved Medical Image Segmentation Through a Deep Learning Paradigm of Multi-Attention Gated Residual U-Net," *IEEE Access*, vol. 12, pp. 40290–40310, 2024, doi: 10.1109/ACCESS.2024.3374108.

[24] J. Nodirov, A. B. Abdusalomov, and T. K. Whangbo, "Attention 3D U-Net with Multiple Skip Connections for Segmentation of Brain Tumor Images," *Sensors*, vol. 22, no. 17, Art. no. 17, Jan. 2022, doi: 10.3390/s22176501.

[25] R. Dwivedi *et al.*, "Explainable AI (XAI): Core Ideas, Techniques, and Solutions," *ACM Comput. Surv.*, vol. 55, no. 9, pp. 1–33, Sep. 2023, doi: 10.1145/3561048.

[26] M. Lucas, M. Lerma, J. Furst, and D. Raicu, "RSI-Grad-CAM: Visual Explanations from Deep Networks via Riemann-Stieltjes Integrated Gradient-Based Localization," in *Advances in Visual Computing*, G. Bebis, B. Li, A. Yao, Y. Liu, Y. Duan, M. Lau, R. Khadka, A. Crisan, and R. Chang, Eds., Cham: Springer International Publishing, 2022, pp. 262–274. doi: 10.1007/978-3-031-20713-6_20.