

Comparative analysis of u-net architectures and variants for hand gesture segmentation in parkinson's patients

Avadhoot Ramgonda Telepatil, Jayashree Sathyanarayana Vaddin

DKTE'S Textile and Engineering Institute Ichalkaranji, Maharashtra, India

Article Info

Article history:

Received Oct 13, 2024

Revised Jun 21, 2025

Accepted Jul 1, 2025

Keywords:

Autoencoder-decoder

Deep learning

Parkinson's disease patients

Segmentation of hand gestures

U-Net framework

ABSTRACT

U-Net is a well-known method for image segmentation, and has proven effective for a variety of segmentation challenges. A deep learning architecture for segmenting hand gestures in parkinson's disease is explored in this paper. We prepared and compared four custom models: a simple U-Net, a three-layer U-Net, an auto encoder-decoder architecture, and a U-Net with dense skip pathways, using a custom dataset of 1,000 hand gesture images and their corresponding masks. Our primary goal was to achieve accurate segmentation of parkinsonian hand gestures, which is crucial for automated diagnosis and monitoring in healthcare. Using metrics including accuracy, precision, recall, intersection over union (IoU), and dice score, we demonstrated that our architectures were effective in delineating hand gestures under different conditions. We also compared the performance of our custom models against pretrained deep learning architectures such as ResNet and VGGNet. Our findings indicate that the custom models effectively address the segmentation task, showcasing promising potential for practical applications in medical diagnostics and healthcare. This work highlights the versatility of our architectures in tackling the unique segmentation challenges associated with parkinson's disease research and clinical practice.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Avadhoot Ramgonda Telepatil

DKTE'S Textile and Engineering Institute Ichalkaranji

Maharashtra-416115, India

Email: avadhootrtelepatil@gmail.com

1. INTRODUCTION

Parkinson's disease affects millions globally, significantly impacting motor skills and quality of life. With 10 million individuals affected by parkinson's disease, innovative solutions are needed to manage symptoms and improve independence. Our research leverages deep learning, particularly U-Net, for accurate hand gesture segmentation, aiding therapeutic interventions and disease monitoring. The U-Net architecture, proposed in [1] has two primary pathways: the contracting path (encoder) and the expansion path (decoder). Similar to a convolutional network, the contracting path extracts classification information while reducing spatial dimensions. Alternatively, the expansion path employs up-convolutions and concatenates features from the contracting path, increasing spatial resolution and enabling the network to refine classification details. U-Net is renowned for its robustness in biomedical image segmentation, especially when dealing with limited annotated data. Enhancements to the U-Net include Swin-U-Net, introduced by Cao *et al.* [2], which integrates swin transformers for improved contextual information extraction, demonstrating superior performance in multi-organ and cardiac segmentation compared to traditional models. A comprehensive review by Siddique *et al.* [3] highlights U-Net's significant impact on medical imaging and notes a growing volume of research output since 2017. Numerous studies have explored

modifications to U-Net for various applications. For instance, Ghaznavi *et al.* [4] compared U-Net variants for segmenting inland water bodies, while Anand *et al.* [5] enhanced the segmentation of dermoscopic skin lesions with a modified architecture, achieving high accuracy. The reduced U-Net, developed by Arun *et al.* [6], maintains high accuracy while simplifying model complexity. Rehman *et al.* [7] created BU-Net for brain tumour segmentation, achieving notable accuracy improvements, and Lu *et al.* [8] introduced half-UNet, which significantly reduces computational requirements while retaining accuracy. These findings underscore U-Net's adaptability and effectiveness across diverse applications, including diabetic retinopathy [9], road segmentation [10], liver image segmentation [11], and dental imaging [12]. UNet-based segmentation achieved high accuracy on the Fish4Knowledge dataset, with potential for further improvement using advanced feature extractors [13]. Newer models such as UNet++ [14] and Res-UNet [15] have refined segmentation capabilities further by enhancing feature fusion and incorporating attention mechanisms. Recent studies demonstrate U-Net's applicability in hand gesture recognition (HGR), particularly for patients with parkinson's disease (PDP). Sreekumar and Geetha [16] evaluated U-Net for hand segmentation in complex backgrounds, achieving an impressive accuracy of 98% on the Egohands dataset and 90% on the GTEA dataset. Similarly, Dutta *et al.* [17] integrated U-Net with VGG16 for HGR, attaining a remarkable recognition rate of 98.97%, demonstrating the model's effectiveness in identifying gestures across multiple classes. Mohan *et al.* [18] further explored various U-Net variants for underwater image segmentation, with VGG-UNet achieving over 98% accuracy in identifying regions of interest (RoI), highlighting its ability to extract significant features from challenging scenes. Pu *et al.* [19] enhanced U-Net for chest X-ray segmentation, achieving a Dice coefficient of 0.973 and 0.983 accuracy, marking a significant advancement in the analysis of X-ray images. Huang and Wang [20] introduced DFP-UNet for brain tumor segmentation, utilizing DenseNet121 to achieve improved accuracy.

Baruah *et al.* [21] employ an attention-based U-Net for brain tumor segmentation, effectively addressing gliomas' heterogeneity using the BraTS dataset. Similarly, Kumar *et al.* [22] demonstrate U-Net's success in lung nodule segmentation using CT images from the LIDC-IDRI dataset, achieving high accuracy and dice similarity coefficient (DSC) scores. Jing *et al.* [23] introduce Mobile-UNet for fabric defect detection, leveraging a lightweight architecture and median frequency balancing to address challenges like data imbalance and computational efficiency. These studies build upon the foundation laid by Shelhamer *et al.* [24], whose fully convolutional network approach set benchmarks in semantic segmentation and inspired numerous advancements in the field. In the realm of patient care, Bernardini *et al.* [25] developed a mobile app for remote monitoring of parkinson's disease patients (PDP), enabling real-time status reporting and intervention. Cardenas *et al.* [26] presented AutoHealth, an IoMT system that employs wearables and AI chatbots for continuous monitoring and personalized care solutions. Ijaz *et al.* [27] focused on lightweight architectures like Mobile U-Net and EfficientU-Net for embedded skin lesion segmentation, delivering improved performance suitable for resource-constrained environments. Lastly, Popat *et al.* [28] enhanced U-Net with additional upsampling boxes in Box-U-Net, improving segmentation metrics such as the Jaccard coefficient. Akter *et al.* [29] enhance U-Net for skin lesion segmentation by integrating a pre-trained Xception encoder. This approach achieves 93.39% accuracy and a dice coefficient of 90.56%, showcasing improved diagnostic performance through transfer learning. The subsequent sections will explore the theoretical foundations of hand gesture segmentation within a PDP assistance system, detailing the methodology for developing the segmentation system, including dataset preparation and U-Net model configurations. Following the discussion of experimental validation, we will examine the results from training U-Net models on selected datasets. Finally, insights will be provided on the study's findings, along with proposed future directions for advancing hand gesture segmentation within PDP systems, highlighting areas for further research and development.

2. BACKGROUND THEORY

Parkinson's disease is a neurological disorder [30]. A patient with involuntary movements like tremors, stiffness, and balance and coordination challenges often needs the assistance of caregivers to manage daily tasks [31]. Caregivers may find it difficult to monitor these patients continuously. While various technologies, including wearable devices that monitor activities and symptoms, have been created to support individuals with PD, many existing assistive technologies do not fully meet the unique needs of these patients. In order to resolve these issues, a vision-based system utilizing HGR through deep learning methods is urgently needed. Effective segmentation of hand gestures is crucial as it improves the accuracy and reliability of subsequent stages, such as gesture classification. This approach aims to enhance the quality of care and support provided to individuals with PD. In this paper, we leverage the adaptable U-Net architecture to investigate multiple customized variants designed for precise segmentation of hand gestures affected by parkinson's disease. These variants include:

2.1. Simple U-Net architecture

The field of image segmentation has witnessed remarkable progress with the introduction of U-Net and its numerous architectural variants. Originally developed for biomedical image segmentation, U-Net's encoder-decoder framework with skip connections has become a foundation for diverse segmentation tasks.

2.2. U-Net with modified layers

An enhanced version of U-Net is features with 3 layers architecture. Three convolutional layers per block, with ReLU activation and batch normalization in encoder. The bottleneck for advanced feature extraction. The decoder Mirrors the encoder structure with three layers per block, concatenating features from the encoder. Finally, the 1×1 convolution produces the segmentation map.

2.3. Encoder-decoder architecture

An architecture focusing on robust feature extraction through an encoder-decoder framework, optimized for capturing subtle variations in hand gestures Starting with a convolutional layer that reduces the input to $256 \times 256 \times 32$, it uses max pooling layers to downsample to $32 \times 32 \times 256$. The decoder restores dimensions back to $256 \times 256 \times 32$, culminating in a final convolution with a sigmoid activation function to produce a $256 \times 256 \times 1$ output.

2.4. U-Net with dense skip pathways

Incorporates dense skip connections to facilitate information flow across different scales of feature maps, improving hand gesture localization accuracy. The advanced U-Net variant uses dense blocks to improve the feature propagation and capture fine details. Its encoder-decoder structure with dense connections enhances segmentation accuracy with a 1×1 final output layer generating the final mask.

A custom dataset of hand gesture images and ground truth masks is used to train and evaluate each modified U-Net architecture. Accuracy, precision, recall, intersection over union (IoU), and dice coefficient are employed to assess segmentation performance and model effectiveness. Through this research, we aim to demonstrate the effectiveness of these modified U-Net architectures in enhancing the segmentation accuracy of parkinsonian hand gestures. In particular, deep learning may improve diagnosis and treatment of neurodegenerative diseases like parkinson's by improving medical image analysis. Contracting and expansive paths are the two main components of the U-Net architecture. CNN blocks are employed in the contracting path, each consisting of 2 consecutive 3×3 convolutions followed by ReLU activation and maximum pooling. This sequence is repeated multiple times to extract features effectively. The innovative aspect of U-Net lies in its expansive path, where each stage involves upsampling the feature map using 2×2 up-convolutions. Following the cropping and concatenation of the contracting map, the upsampled map is convoluted two times with ReLU activation. Finally, a 1×1 convolution reduces the feature map to the desired number of channels for segmentation output. The cropping step is crucial as it eliminates edge pixels with minimal contextual information, resulting in a U-shaped network structure. This design facilitates the propagation of contextual information across the network, enabling effective object segmentation by utilizing context from a broader surrounding area. The network's energy function is given by (1) and (2),

$$E = \sum w(x) \log(p_{k(x)}(x)) \quad (1)$$

$$p_k = \frac{\exp(a_k(x))}{\sum_{k'}^k \exp(a_k(x)')} \quad (2)$$

here p_k represents the pixel-wise SoftMax function applied over the final feature map, a_k denotes the activation in channel k .

3. METHOD

3.1. Dataset preparation

Due to the unavailability of suitable public datasets for HGR in PDP, we formulated the dataset which includes images of hand gestures that are specifically formulated for ease of execution by PDP. Each gesture image is paired with an equivalent ground truth mask, which provides pixel-wise annotations for the performed gesture. Each image in the dataset is grayscale and has a resolution of 256×256 pixels. Examples of these images and their corresponding masks are shown in Table 1.

3.1.1. Hand gesture images

The practical and accessible dataset play a vital role in segmentation activity. The dataset was captured using a standard USB webcam to support a low-cost and accessible setup. Images were recorded



















under consistent lighting and background conditions to ensure data uniformity. This approach helps reduce variability and improves the reliability of segmentation results. Such a setup also reflects real-world deployment scenarios.

The practical and accessible dataset play a vital role in segmentation activity. The dataset was captured using a standard USB webcam to support a low-cost and accessible setup. Images were recorded under consistent lighting and background conditions to ensure data uniformity. This approach helps reduce variability and improves the reliability of segmentation results. Such a setup also reflects real-world deployment scenarios.

3.1.2. Ground truth masks

Binary masks corresponding to each hand gesture image, where gesture regions are labelled distinctly from the background. Hand gesture images were recorded using a USB camera to ensure both accessibility and practicality. This approach provided real-time, high-resolution images suitable for our segmentation models. We made efforts to standardize the image quality and lighting conditions to minimize variability.

Table 1. Independent data set of hand gesture with meaning and posture with ground truth masks

Sr. No.	Posture	Ground truth mask	Meaning
1			Indicates the need for water
2			Signifies hunger
3			Represents a natural call
4			need to use the toilet
5			Calls for attention
6			Indicates a desire to move
7			Signifies turning the TV on or off
8			Represents agreement
9			Indicates happiness

3.2. Model architecture

This section outlines the models implemented for hand gesture segmentation in PDP. It describes the layered architecture implemented for the segmentation operation. It first deals with practical details and implementation of the standard U-Net architecture. Followed to this, the modifications in standard U-Net such as variation in layers, the implementation of the encoder decoder structure is explored.

3.2.1. U-Net architecture

The basic U-Net model is designed for processing grayscale images of size 256×256 and includes:

- Encoder: composed of convolutional layers with max pooling for down sampling. Each block consists of two 3×3 convolutions with ReLU activation and batch normalization.
- Bottleneck: contains convolutional layers for deep feature extraction.
- Decoder: utilizes transposed convolutions for Upsampling and incorporates skip connections from the encoder to enhance segmentation.
- Output layer: A 1×1 convolution generates the final segmentation mask.

The contracting path captures high-level features through increasing filter counts (16, 32, 64, 128, 256), while the expansive path up samples features maps and merges them with corresponding encoder features. This process culminates in a single-channel output representing the segmented result.

3.2.2. U-Net with 3 layers

An enhanced version of U-Net features:

- Encoder: three convolutional layers per block, with ReLU activation and batch normalization.
- Bottleneck: three convolutional layers for advanced feature extraction.
- Decoder: mirrors the encoder structure with three layers per block, concatenating features from the encoder.
- Output layer: a 1×1 convolution produces the segmentation map.

This model processes grayscale images of size 256×256, with filter counts progressively increasing from 64 to 256. The decoder reconstructs image dimensions back to 256×256, using a final convolutional layer with a single filter and sigmoid activation to produce a 256×256×1 output.

3.2.3. Encoder-decoder model

This convolutional autoencoder model also processes grayscale images of size 256×256:

- Encoder: a series of convolutional layers for down sampling and feature extraction.
- Bottleneck: additional convolutional layers for feature processing.
- Decoder: employs upsampling layers and skip connections to reconstruct the image.
- Output layer: a final convolutional layer generates the segmentation mask.

Starting with a convolutional layer that reduces the input to 256×256×32, it uses max pooling layers to downsample to 32×32×256. The decoder restores dimensions back to 256×256×32, culminating in a final convolution with a sigmoid activation function to produce a 256×256×1 output.

3.2.4. U-Net with dense skip pathway

This advanced U-Net variant enhances feature propagation and segmentation accuracy:

- Encoder: contains dense blocks in which each convolutional layer receives input from its predecessors.
- Dense blocks: enhance feature propagation and capture intricate details.
- Decoder: utilizes dense connections and concatenates features from the encoder.
- Output layer: a 1×1 convolution produces the final segmentation mask.

The encoder captures hierarchical features through four convolutional blocks (64 to 512 filters) with max pooling, while the decoder uses upsampling and skip connections to refine gesture segmentation. These models enhance segmentation accuracy for PDP, aiding in better therapy and monitoring.

3.3. Evaluation metrics

Different evaluation metrics are used to assess the performance of image segmentation models. Each metric provides insights into different aspects of the segmentation quality. The following metrics were used:

3.3.1. Dice coefficient

The dice coefficient, additionally known as the DSC, measures the overlap among the expected segmentation masks and the ground truth mask. It is beneficial for evaluating the overall performance in situations with imbalanced training. Mathematically is represented as,

$$\text{Dice Coefficient} = \frac{2 \cdot |A \cap B|}{|A| + |B|} \quad (3)$$

where,

- $|A \cap B|$ = Number of pixels in the intersection of the predicted mask A and the ground truth mask B
- $|A|$ = Number of pixels in the predicted mask A and $|B|$ = Number of pixels in the ground truth mask B

The range of these metrics is from 0 to 1, where a value of 0 indicates no overlap or accuracy, and a value of 1 represents perfect overlap or complete accuracy.

3.3.2. IoU

IoU, also known as the Jaccard Index, evaluates the quality of the segmentation by comparing the intersection and union of the predicted and ground truth masks. It is widely used for its simplicity and effectiveness in evaluating segmentation quality.

$$\text{Intersection over Union} = \frac{|A \cap B|}{|A \cup B|} \quad (4)$$

Where,

- $|A \cap B|$ = Number of pixels in the intersection of the predicted mask A and the ground truth mask B
- $|A \cup B|$ = Number of pixels in the union of the predicted mask A and the ground truth mask B

The metrics range from 0 to 1, where 0 signifies no overlap and 1 denotes perfect overlap.

3.3.3. Pixel accuracy

Pixel accuracy is a measure of how many pixels are correctly classified out of all the pixels in the image. By measuring the performance of the model across the entire image, it can be used to assess how well the model is performing. Pixel accuracy is a measure of how many pixels are correctly classified from the total number of pixels. An overall measure of the model's performance across the entire image is provided by it.

$$\text{Pixel Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

3.3.4. Precision and recall

Precision and recall provide deeper insights into the model's accuracy and its ability to detect positive instances. Precision measures the proportion of true positive predictions among all positive predictions made by the model. It is useful for understanding the accuracy of positive classifications.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (6)$$

Recall measures the proportion of true positive pixels among all actual positive pixels. It reflects the model's ability to detect positive instances.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (7)$$

Where, TP is the number of correctly classified foreground pixels. TN is number of correctly classified background pixels, FP is number of backgrounds incorrectly classified pixels and FN is incorrectly classified background pixels. These metrics provide a comprehensive view of the model's performance in segmenting images, helping to evaluate both the quality of the segmentation and the model's effectiveness in distinguishing between different classes.

3.4. Experimentation

The experimental steps followed in the segmentation process for the parkinson's disease patients hand gesture (PDP-HG) segmentation system are outlined as follows:

3.4.1. Dataset preparation

A dataset of PDP hand gestures was created, simulating potential neurological deficits. This dataset included 1000 hand gesture images. These images are categorized into 9 different classes and are as shown in Table 1. This dataset includes original hand gesture images and their corresponding segmentation masks.

3.4.2. Dataset splitting

The dataset splitting is very important in deep learning to evaluate how well the model works to new, unseen data. It is helpful for model to learn the patterns, fine tuning to work effieciently. The dataset was split into training, testing, and validation subsets to ensure effective model training and evaluation. This partitioning helps assess U-Net variants on unseen data for reliable performance.

3.4.3. Model configuration

The U-Net model and its variants-including U-Net, U-Net with 3 layers, encoder-decoder, and U-Net with dense skip connections-were configured. This involved defining the network architecture and implementing specific modifications tailored to the segmentation task.

3.4.4. Performance metrics calculation

Performance metrics, as detailed in section 3.3., were computed, including accuracy, recall, precision, IoU, and dice score. A comprehensive evaluation of the models' capability to identify and delineate hand gestures is provided by these metrics.

3.4.5. Model comparison with pretrained models

Pretrained models, such as ResNet and VGGNet, were applied to the dataset. The results from these models were used for comparison with the performance of the implemented models. Hand gesture segmentation for parkinson's patients was implemented using U-Net and its variants on Google Colab with an NVIDIA T4 GPU, leveraging a custom dataset.

4. RESULT AND DISCUSSION

The U-Net model and its various variants were evaluated on a self-generated hand gesture image dataset, as outlined in section 3.1. To train the models, 32 batches were divided into 50 epochs. The model accuracy and model loss are important parameters in deep learning. The performance of the U-Net architecture is illustrated in Figure 1. It presents the overall performance of the U-Net architecture during training and validation phases in form of model accuracy and model loss. It shows these parameters changes over 50 epochs. Specially Figure 1(a) the accuracy trend and Figure 1(b) the loss progression. From the graphs, it is observed that the training loss achieved was 0.20, with a corresponding training accuracy of 0.90. On the validation dataset, the model yielded a validation loss of 0.2717 and a validation accuracy of 0.8756.

The performance of modified U-Net architecture in form of the model's accuracy and loss throughout the training and validation phases across various 50 epochs is shown in Figure 2. Further, the performance metrics of the modified U-Net architecture are illustrated in Figures 2(a) and 2(b). The results indicate that the model achieved a training loss of 0.2853 and an accuracy of 0.8629, while the validation loss was 0.2649 with a corresponding accuracy of 0.8741. The encoder decoder model performance in form of correct predictions of segmentation mask along with the errors between models predicted output and actual target value in the form of model accuracy and model loss is illustrated in Figure 3. As shown in Figures 3(a) and 3(b), the model achieved a training loss of 0.0198 and an accuracy of 0.9541, while the validation loss was 0.0371 with a validation accuracy of 0.9328.

Figure 4 illustrates the overall performance of the modified U-Net architecture with dense skip connections during training and validation across 50 epochs. It provides a comprehensive view of how the model's accuracy and loss evolved throughout the learning process. As shown in Figure 4(a) and 4(b), the model attained a training loss of 0.2403 and an accuracy of 0.8895, with a validation loss of 0.2076 and a validation accuracy of 0.9055. Table 2 displays the performance metrics for each model, including accuracy, precision, recall, IoU, and dice score based on the testing dataset.

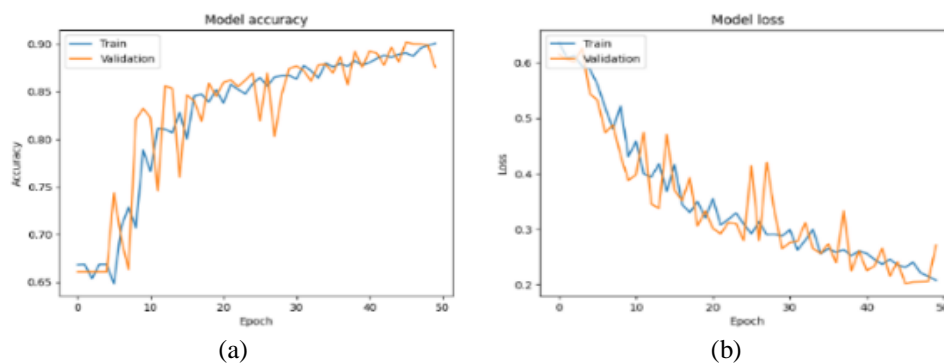


Figure 1. Model performance for U-Net architecture over the training and validation datasets across the epochs (a) model accuracy and (b) model loss

Table 2. Performance evaluation of different segmentation models

Model name	Evaluation parameters				
	Accuracy	Precision	Recall	IoU	Dice
U-Net architecture	93.61%	0.88	0.92	0.82	0.9
Modified U-Net architecture	89.49%	0.83	0.85	0.72	0.83
Autoencoder decoder architecture	95.71%	0.94	0.92	0.87	0.93
U-Net with dense skip connections	92.88%	0.89	0.89	0.8	0.88
Pretrained model with ResNet	88.13%	0.82	0.8	0.69	0.81
Pretrained model with VGGNet	94.62%	0.9	0.91	0.85	0.91

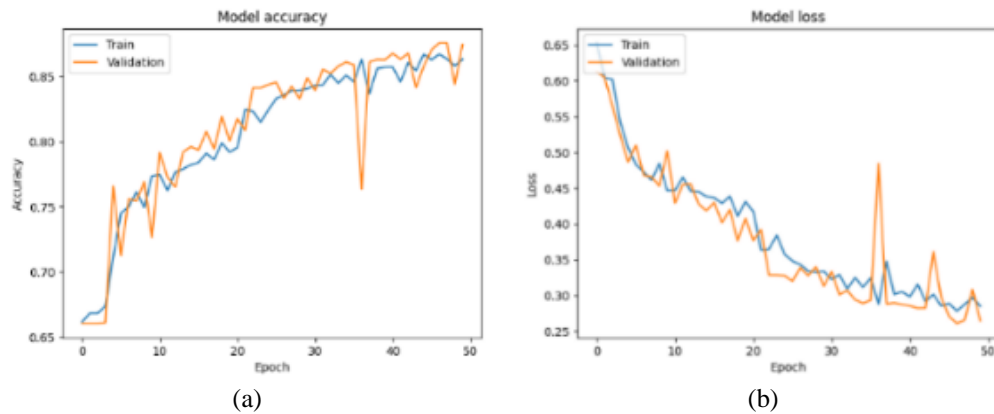


Figure 2. Model performance for modified U-Net architecture over the training and validation datasets across the epochs (a) model accuracy and (b) model loss

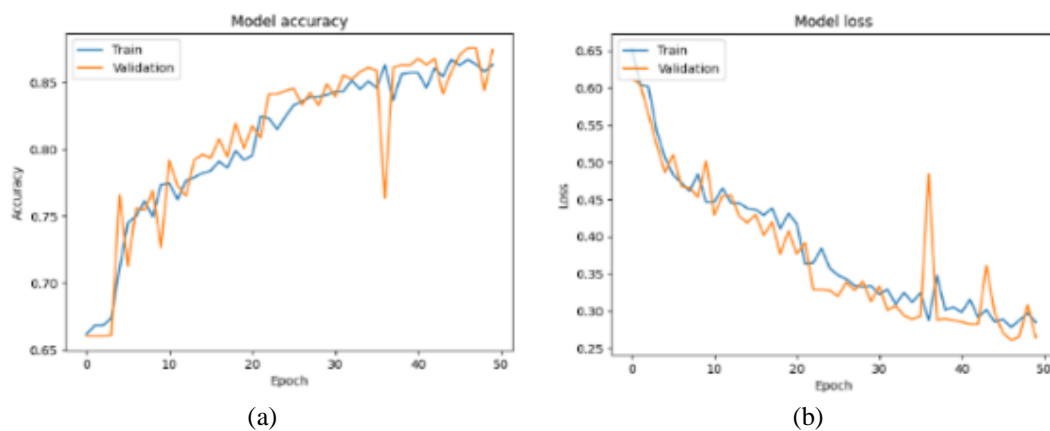


Figure 3. Model performance for encoder decoder architecture over the training and validation datasets across the epochs (a) model accuracy and (b) model loss

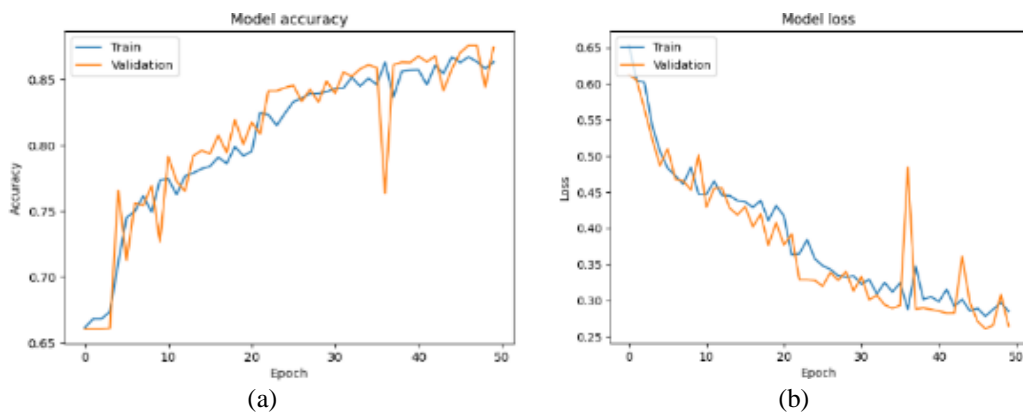


Figure 4. Model performance for modified U-Net architecture with dense skip connections architecture over the training and validation datasets across the epochs (a) model accuracy and (b) model loss

The autoencoders decoder architecture outperforms all other models across key metrics, demonstrating the highest overall effectiveness in segmentation tasks. With an accuracy of 95.71%, it excels in correctly classifying pixels and achieves the highest precision (0.94), recall (0.92), IoU (0.87), and dice score (0.93). This indicates exceptional performance in both identifying and delineating object pixels. The standard U-Net architecture follows closely with a 93.61% accuracy, high precision (0.88), recall (0.92), IoU (0.82), and dice score (0.90), showing strong performance but slightly behind the autoencoder decoder.

The Modified U-Net architecture performs the least favorably, with an accuracy of 89.49%, and lower scores across precision (0.83), recall (0.85), IoU (0.72), and dice (0.83), suggesting that its modifications may have reduced its overall effectiveness. The U-Net with dense skip connections shows robust performance with a 92.88% accuracy, good precision (0.89), recall (0.89), IoU (0.80), and dice score (0.88). Although it benefits from dense skip connections, it does not match the autoencoder decoder’s performance. Overall, the autoencoder decoder architecture is the most effective model for segmentation in this experiment, providing superior results in accuracy, precision, recall, IoU, and dice score.

Custom segmentation models outperform pretrained ResNet and VGGNet in accurately segmenting hand gestures for parkinson’s patients. While the autoencoder-decoder architecture achieved the highest accuracy at 95.71%, surpassing both pretrained models, the standard U-Net also demonstrated robust performance with an accuracy of 93.61%. In contrast, the pretrained ResNet model underperformed with an accuracy of 88.13%, indicating challenges in effectively identifying and segmenting gestures. The VGGNet model fared better, achieving an accuracy of 94.62%, but still fell short compared to the best-performing custom architectures. This suggests that while pretrained models can provide a solid foundation, custom-designed architectures are better suited for the specific hand gesture segmentation in this clinical context. The results are organized into a comprehensive figure as shown in Figure 6. The first model assessed is the standard U-Net, next we examined a modified U-Net with three layers following this, we analyzed an encoder-decoder model, lastly a U-Net variant with dense skip connections, an advanced architecture that integrates dense connections between the encoder and decoder. Each row in the table provides a comparative view of the ground truth masks alongside the predicted masks from these models, highlighting the effectiveness and performance differences across the different architectural approaches. This structured comparison allows for a clear assessment of how each model variant performs relative to the others and the ground truth.

























Model	Ground Truth Mask	Predicted Mask	Model	Ground Truth Mask	Predicted Mask	Model	Ground Truth Mask	Predicted Mask	Model	Ground Truth Mask	Predicted Mask
UNET			Modified UNET			Encoder Decoder			UNET with Dense Skip Connections		
											
											

Figure 6. Comparison of ground truth and predicted masks for various deep learning models

5. CONCLUSION

The aim of this study was to identify the most accurate and robust hand gesture segmentation architecture for PDP. We developed four models: a standard U-Net, a three-layer U-Net variant, an autoencoder-decoder architecture, and a U-Net with dense skip pathways, all trained on a custom dataset of 1,000 images and corresponding masks. The autoencoder-decoder architecture emerged as the most effective, achieving an accuracy of 95.71% and high precision (0.94), recall (0.92), IoU (0.87), and dice score (0.93). The standard U-Net also performed well with an accuracy of 93.61%. In contrast, the modified U-Net showed lower performance (89.49%), indicating that its enhancements did not yield the desired improvements. When compared to pretrained models like ResNet and VGGNet, our custom architectures outperformed these options, particularly highlighting the tailored models’ strengths in this clinical context. The results underscore the potential of advanced deep learning architectures to enhance medical image analysis and improve diagnostic capabilities in neurodegenerative diseases. Future research may involve refining these models and exploring additional architectures to further advance automated gesture recognition for parkinson’s disease.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Avadhoot Ramgonda Telepatil	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	
Jayashree Sathyanarayana Vaddin	✓	✓			✓	✓	✓	✓	✓	✓	✓	✓		

C : Conceptualization

M : Methodology

So : Software

Va : Validation

Fo : Formal analysis

I : Investigation

R : Resources

D : Data Curation

O : Writing - Original Draft

E : Writing - Review & Editing

Vi : Visualization

Su : Supervision

P : Project administration

Fu : Funding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The dataset used in this study was self-generated using a USB webcam under controlled conditions and is not publicly available due to privacy and project-specific constraints. However, the data can be made available from the corresponding author upon reasonable request for academic and research purposes.




REFERENCES

- [1] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9351, pp. 234–241, 2015, doi: 10.1007/978-3-319-24574-4_28.
- [2] H. Cao *et al.*, "Swin-Unet: Unet-like pure transformer for medical image segmentation," in *Lecture Notes in Computer Science*, vol. 13803 LNCS, 2023, pp. 205–218.
- [3] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, "U-Net and its variants for medical image segmentation: a review of theory and applications," *IEEE Access*, vol. 9, pp. 82031–82057, 2021, doi: 10.1109/ACCESS.2021.3086020.
- [4] A. Ghaznavi, M. Saberioon, J. Brom, and S. Itzerott, "Comparative performance analysis of simple U-Net, residual attention U-Net, and VGG16-U-Net for inventory inland water bodies," *Applied Computing and Geosciences*, vol. 21, p. 100150, Mar. 2024, doi: 10.1016/j.acags.2023.100150.
- [5] V. Anand, S. Gupta, D. Koundal, S. R. Nayak, P. Barsocchi, and A. K. Bhoi, "Modified U-NET architecture for segmentation of skin lesion," *Sensors*, vol. 22, no. 3, p. 867, Jan. 2022, doi: 10.3390/s22030867.
- [6] R. A. Arun, S. Umamaheswari, and A. V. Jain, "Reduced U-Net architecture for classifying crop and weed using pixel-wise segmentation," in *2020 IEEE International Conference for Innovation in Technology, INOCON 2020*, Nov. 2020, pp. 1–6, doi: 10.1109/INOCON50539.2020.9298209.
- [7] M. U. Rehman, S. Cho, J. H. Kim, and K. T. Chong, "Bu-net: brain tumor segmentation using modified u-net architecture," *Electronics (Switzerland)*, vol. 9, no. 12, pp. 1–12, Dec. 2020, doi: 10.3390/electronics9122203.
- [8] H. Lu, Y. She, J. Tie, and S. Xu, "Half-UNet: a simplified U-Net architecture for medical image segmentation," *Frontiers in Neuroinformatics*, vol. 16, Jun. 2022, doi: 10.3389/fninf.2022.911679.
- [9] N. Sambyal, P. Saini, R. Syal, and V. Gupta, "Modified U-Net architecture for semantic segmentation of diabetic retinopathy images," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 3, pp. 1094–1109, Jul. 2020, doi: 10.1016/j.bbe.2020.05.006.
- [10] N. Y. Q. Abderrahim, S. Abderrahim, and A. Rida, "Road segmentation using u-net architecture," in *Proceedings - 2020 IEEE International Conference of Moroccan Geomatics, MORGeo 2020*, May 2020, pp. 1–4, doi: 10.1109/MorGeo49228.2020.9121887.
- [11] X. Li, W. Qian, D. Xu, and C. Liu, "Image segmentation based on improved Unet," *Journal of Physics: Conference Series*, vol. 1815, no. 1, p. 012018, Feb. 2021, doi: 10.1088/1742-6596/1815/1/012018.
- [12] S. Sivagami, P. Chitra, G. S. R. Kailash, and S. R. Muralidharan, "UNet architecture based dental panoramic image segmentation," in *2020 International Conference on Wireless Communications, Signal Processing and Networking, WiSPNET 2020*, Aug. 2020, pp. 187–191, doi: 10.1109/WiSPNET48689.2020.9198370.
- [13] N. A. Nezla, T. P. Mithun Haridas, and M. H. Supriya, "Semantic segmentation of underwater images using UNet architecture based deep convolutional encoder decoder model," in *2021 7th International Conference on Advanced Computing and Communication Systems, ICACCS 2021*, Mar. 2021, pp. 28–33, doi: 10.1109/ICACCS51430.2021.9441804.
- [14] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 1856–1867, Jun. 2020, doi: 10.1109/TMI.2019.2959609.
- [15] X. Li, Z. Fang, R. Zhao, and H. Mo, "Brain tumor MRI segmentation method based on improved Res-UNet," *IEEE Journal of Radio Frequency Identification*, vol. 8, pp. 652–657, 2024, doi: 10.1109/JRFID.2023.3349193.
- [16] A. Sreekumar and M. Geetha, "Hand segmentation in complex background using UNet," in *Proceedings of the 2nd International Conference on Inventive Research in Computing Applications, ICIRCA 2020*, Jul. 2020, pp. 440–445, doi: 10.1109/ICIRCA48905.2020.9183215.
- [17] H. P. J. Dutta, D. Sarma, M. K. Bhuyan, and R. H. Laskar, "Semantic segmentation based hand gesture recognition using deep neural networks," in *26th National Conference on Communications, NCC 2020*, Feb. 2020, pp. 1–6, doi: 10.1109/NCC48643.2020.9055990.




- [18] R. Mohan, M. Abouhawwash, R. Arunmozhi, and V. Rajinikanth, "Automatic segmentation of underwater images with shannon's thresholding and UNet variants," in *Winter Summit on Smart Computing and Networks, WiSSCoN 2023*, Mar. 2023, pp. 1–6, doi: 10.1109/WiSSCoN56857.2023.10133852.
- [19] Q. Pu, D. Wei, and J. Tian, "Chest X-ray image segmentation based on improved UNet," in *2023 8th International Conference on Intelligent Computing and Signal Processing, ICSP 2023*, Apr. 2023, pp. 1842–1846, doi: 10.1109/ICSP58490.2023.10248884.
- [20] W. Huang and J. Wang, "Automatic segmentation of brain tumors based on DFP-UNet," in *IEEE 6th Information Technology and Mechatronics Engineering Conference, ITOEC 2022*, Mar. 2022, pp. 1304–1307, doi: 10.1109/ITOEC53115.2022.9734456.
- [21] P. Baruah, B. Dutta, P. P. Dutta, H. Pallab Jyoti Dutta, B. Goswami, and D. Sarma, "Brain tumor segmentation using attention-based UNet," in *Proceedings of 2023 IEEE 3rd Applied Signal Processing Conference, ASPCON 2023*, Nov. 2023, pp. 282–286, doi: 10.1109/ASPCON59071.2023.10396058.
- [22] S. N. Kumar *et al.*, "Lung nodule segmentation using UNet," in *2021 7th International Conference on Advanced Computing and Communication Systems, ICACCS 2021*, Mar. 2021, pp. 420–424, doi: 10.1109/ICACCS51430.2021.9441977.
- [23] J. Jing, Z. Wang, M. Rättsch, and H. Zhang, "Mobile-Unet: an efficient convolutional neural network for fabric defect detection," *Textile Research Journal*, vol. 92, no. 1–2, pp. 30–42, Jan. 2022, doi: 10.1177/0040517520928604.
- [24] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 640–651, Apr. 2017, doi: 10.1109/TPAMI.2016.2572683.
- [25] S. Bernardini, C. Cianfrocca, M. Maioni, M. Pennacchini, D. Tartaglioni, and L. Vollero, "A Mobile App for the remote monitoring and assistance of patients with parkinson's disease and their caregivers," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, Jul. 2018, vol. 2018-July, pp. 2909–2912, doi: 10.1109/EMBC.2018.8512989.
- [26] L. Cardenas, K. Parajes, M. Zhu, and S. Zhai, "AutoHealth: advanced LLM-empowered wearable personalized medical butler for parkinson's disease management," in *2024 IEEE 14th Annual Computing and Communication Workshop and Conference, CCWC 2024*, Jan. 2024, pp. 375–379, doi: 10.1109/CCWC60891.2024.10427622.
- [27] H. Ijaz, H. Sultan, M. Altaf, and A. Waris, "Embedded skin lesion segmentation using lightweight encoder-decoder architectures," in *3rd IEEE International Conference on Artificial Intelligence, ICAI 2023*, Feb. 2023, pp. 176–181, doi: 10.1109/ICAIS8407.2023.10136688.
- [28] M. Popat, S. Patel, Y. Poshia, and A. K. Sai, "Brain tumor image segmentation using Box-Unet architecture," in *2023 1st International Conference on Advances in Electrical, Electronics and Computational Intelligence, ICAEECI 2023*, Oct. 2023, pp. 1–7, doi: 10.1109/ICAEECI58247.2023.10370887.
- [29] A. Akter, K. Deb, S. C. Tista, and K. H. Jo, "A modified UNet for skin lesion segmentation using transfer learning," in *Proceedings - IWIS 2023: 3rd International Workshop on Intelligent Systems*, Aug. 2023, pp. 1–6, doi: 10.1109/IWIS58789.2023.10284642.
- [30] A. Telepatil and J. Vaddin, "Various optimizers' performances for CNN-based hand gesture recognition for PDP assistance," in *2023 14th International Conference on Computing Communication and Networking Technologies, ICCCNT 2023*, Jul. 2023, pp. 1–7, doi: 10.1109/ICCCNT56998.2023.10307504.
- [31] Banita, "Detection of parkinson's disease using rating scale," in *2020 International Conference on Computational Performance Evaluation, ComPE 2020*, Jul. 2020, pp. 121–125, doi: 10.1109/ComPE49325.2020.9200071.

BIOGRAPHIES OF AUTHORS



Avadhoot Ramgonda Telepatil    completed his Bachelor of Engineering (B.E.) in Electronics and Telecommunication and an M.E. in Electronics Engineering from Shivaji University, Kolhapur. He is currently pursuing a Ph.D. at Shivaji University and serves as an Assistant Professor at D.K.T.E's Textile and Engineering Institute, Ichalkaranji, with 15 years of teaching experience. His research interests include image processing, embedded systems, AI, and ML. He is committed to innovation, mentoring students, and advancing technology. He can be contacted at email: avadhoottelepatil@gmail.com.



Dr. Jayashree Sathyanarayana Vaddin   , Ph.D. (Electronics, 2013), M.E. (Electronics, 1997)–Shivaji University, B.E. (Electrical, 1983)–Karnataka University. Former Professor and HOD, she established seven Electrical Engineering labs and a VLSI design lab under an AICTE grant. With 38 years of teaching experience, her research focuses on image processing, AI, and VLSI design. She has published 70 papers, 4 books, filed one patent, and delivered 18 expert lectures. She has mentored 21 PG and 6 Ph.D. students. A senior IEEE member since 2019, she has received multiple awards. She can be contacted at email: jayashreevaddin@gmail.com.