International Journal of Scientific Research in Computer Science and Engineering
Vol.**13**, Issue.**1**, pp.**26-31**, February **2025**
E-ISSN: 2320-7639
Available online at: www.isroset.org

Research Article

An Empirical Analysis of Encoder-Decoder (U-Net) Variants for Medical Image Segmentation

Jackson Kamiri[1*], Geoffrey Mariga Wambugu[2], Aaron Mogeni Oirere[3]

[1,2]Dept. of Information Technology, Murang'a University of Technology, Murang'a, Kenya
[3]Dept. of Computer Science, Murang'a University of Technology, Murang'a, Kenya

**Corresponding Author:* ✉

Received: 18/Dec/2024; **Accepted**: 20/Jan/2025; **Published**: 28/Feb/ | https://doi.org/10.26438/ijsrcse.v13i1.610

Copyright © 2025 by author(s). This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited & its authors credited.

Abstract— U-Net convolutional neural networks have become a cornerstone in medical image processing, particularly for complex segmentation tasks. However, with the proliferation of various U-Net variants, it is imperative to evaluate their performance across diverse medical datasets to determine the most suitable architecture for specific applications. This study presents a comprehensive empirical analysis of six U-Net variants namely; U-Net, U-Net++, ResU-Net, TransU-Net, V-Net, and U-Net 3+, focusing on their effectiveness in medical image segmentation. The research used two publicly available medical image datasets: the Breast Cancer Ultrasound dataset and the Brain MRI FLAIR dataset. These datasets were selected to represent a range of challenges in medical imaging, including varying levels of noise, contrast, color channels, and lesion visibility. Each U-Net variant was trained and evaluated under consistent experimental conditions, with the Dice Similarity Coefficient (DSC), Jaccard Index and accuracy serving as the primary evaluation metrics. The findings reveal that while all U-Net variants exhibit strong performance, ResU-Net consistently outperformed other architectures across both datasets. Specifically, ResU-Net achieved the highest Jaccard Index and DSC values, recording 0.5158 and 0.5930 for the Breast Cancer Ultrasound dataset and 0.7923 and 0.8730 for the Brain MRI dataset, respectively. Additionally, ResU-Net demonstrated the highest accuracy of 96.20% in the Breast Ultrasound dataset and emerged as the first runner-up in the Brain MRI dataset with an accuracy of 99.70%, trailing slightly behind U-Net++, which achieved 97.72%. Conversely, U-Net 3+ registered the lowest performance in both datasets. ResU-Net's superior performance suggests that its residual connections enhance feature propagation and mitigate the vanishing gradient problem, leading to better segmentation outcomes. Conversely, U-Net 3+ showed lower performance, indicating that complexity doesn't always equate to better results. This study offers insights into the strengths and weaknesses of different U-Net variants, guiding practitioners in choosing the most suitable model for their specific applications especially in medical image processing domain.

Keywords — Medical Image Segmentation, U-Net variants, Brain MRI, Breast Ultrasound, ResU-Net, Encoder-Decoder, Image processing, Computer Vision.

1. Introduction

Medical image segmentation is a pivotal process in the field of medical imaging, where it serves as the foundation for a wide range of clinical applications, from diagnosis to treatment planning and research. This technique involves the partitioning of medical images, such as MRI, CT scans, or ultrasound images, into distinct regions that correspond to different anatomical structures, tissues, or pathological areas[1]. By delineating these regions, segmentation transforms complex medical images into more comprehensible and quantifiable forms, enabling healthcare professionals to identify and analyze specific areas of interest, such as tumors, organs, blood vessels, or lesions, with greater precision[2]. This allows healthcare professionals to make more precise diagnoses, plan surgical interventions, and monitor treatment outcomes with greater confidence.

Medical images are acquired using different modalities with the most popular ones being Magnetic resonance imagining (MRI), X-ray, PET scan, Computed Tomography (CT), Ultrasound, Retinal imaging, Dermoscopy and Histopathology[3]. Some modalities also have sub-modalities within them. Magnetic resonance imaging being the most used technique for brain tumors has several modalities such as "Fluid-Attenuated Inversion Recovery (FLAIR), T1 Weighted (T1W), T1-Weighted with Contrast Enhancement (T1wCE),

and T2-Weighted" [3]. Each of the medical imaging modalities has unique features and varying levels of complexity.

Image segmentation tasks can be grouped into either semantic segmentation or instance segmentation. "Semantic segmentation assigns a class label to each pixel in an image, treating all instances of the same class as a single entity, while instance segmentation not only classifies each pixel but also distinguishes between different instances of the same class within the image"[4]. Traditional techniques, such as thresholding and edge detection, were used previously, but recent advancements in machine learning and deep learning have revolutionized the field, leading to the development of sophisticated models. The emergence of imaging machines that can generate multi-modal images opened a new frontier for researchers to develop deep learning solutions that can perform multi-modal segmentation through image fusion[5], [6]. Among the various segmentation models developed, U-Net and its variants have emerged as some of the most effective and widely used architectures, particularly in tasks that require high accuracy and efficiency[7]. Originally introduced for biomedical image segmentation, the architecture consists of contracting paths and symmetrical expanding paths to capture context and enable precise localization respectively[7].

The U-Net architecture has been adapted and modified to suit a wide range of medical imaging applications. These adaptations, often referred to as U-Net variants, have been specifically designed to address the unique challenges posed by different types of medical images and medical imaging modalities, such as variations in resolution, contrast, and noise. This study aimed at comparing six U-Net variants namely; U-Net, U-Net++, ResU-Net, V-Net, TransU-Net, and U-Net3+. The empirical analysis involved training the architectures within the same environmental setup for each of the dataset in order to achieve fair comparison. Although, other studies such as Ehab and Li [8], have done a similar study, their study only considered three U-Net architectures, therefore, this paper presents a more comprehensive analysis compared to[8]. The findings of this study are critical in guiding researchers in choosing the most optimal model.

2 Literature Review

Convolutional neural networks have demonstrated great performance in image processing [9].The encoder-decoder architecture forms the basis of U-Net convolutional Neural Networks. However, several variants of this architecture have emerged as improvements to the standard U-Net. They include; U-Net++, ResU-Net, TransU-Net, V-Net, and U-Net3+

2.1 U-Net

U-Net was develop by Ronneberger et al., [7] to perform medical image segmentation tasks. The base U-Net consists of two paths; "the contracting path also known as the encoder and the expanding path also known as the decoder"[10]. The contracting path consists of convolutional neural network (CNN) blocks. Each block has 3× 3 convolutions followed by a Rectified Linear Unit (ReLu) activation function and MaxPooling layer. The number of such blocks used depends on the nature of task that the developer is performing. "The innovation of U-Net lies in its second component, known as the expansive path. In this phase, each stage up-samples the feature map using 2 × 2 up-convolutions. The feature map from the corresponding layer in the contracting path is then cropped and concatenated with the up-sampled feature map. This process is followed by two consecutive 3 × 3 convolutions with ReLu activation. In the final stage, a 1 × 1 convolution is applied to reduce the feature map to the desired number of channels, resulting in the segmented image"[10], [11]. Cropping is essential because pixel features at the edges contain minimal contextual information and therefore need to be excluded.

This design forms a U-shaped network that effectively propagates contextual information throughout, enabling the segmentation of objects within a region by leveraging context from a broader overlapping area[7], [10].

U-Net can be represented as follows:

Let X_0 be the input image, each convolution in layer l is represented as; $C_l = RELU(Conv_{3\times3}(X_{l-1}))$, after two convolutions $X_l = RELU(Conv_{3\times3}(C_l))$. The MaxPooling operation can be represented as: $P_l = MaxPool_{2\times2}(X_l)$. The network bottleneck which applies 3×3 convolutions followed by ReLU activation without pooling can be represented as; $B = RELU(Conv_{3\times3}\left(RELU\left(Conv_{3\times3}(P_L)\right)\right))$ where L is the index of the last layer in the contracting path. The upsampling in the expansive path can be represented as; Let U_K represent the up-convolution (transposed convolution), $U_K = UpConv_{2\times2}(B_{K+1})$. Concatenate U_K with the corresponding feature map X_{L-K} from the contracting path to get; $C_K = Concat(U_K, X_{L-K})$. Apply the 3 × 3 convolutions followed by ReLu; $D_K = RELU(Conv_{3\times3}\left(RELU\left(Conv_{3\times3}(C_K)\right)\right))$. The final output layer $Y = Conv_{1\times1}(D_0)$ where D_0 is the feature map from the last layer of the expansive path.

2.2 U-Net ++

Introduced by Zhou et al., [12], it was fronted as an advancement of the basic U-Net where the encoder and decoder subnetworks are connected through a series of nested dense skip pathways[12]. This new architecture improved the learning capabilities of the optimizer since the skip connections reduced the semantic gap between the feature maps of the encoder and the decoder subnetworks. The U-Net++ allows for concatenation to be greatly refined at higher levels of the decoder sub-network[13]. Unlike the standard U-Net where the feature maps in the encoder are directly transmitted to the decoder, the U-Net++ architecture passes the feature maps through some dense convolution blocks. U-Net++ uses deep supervision thus allowing the model to operate in two modes namely; accurate mode which averages outputs from all segmentation branches and the fast mode in which the final segmentation map is selected from only one of the segmentation branches[12].

© 2025, IJSRCSE All Rights Reserved

2.3 ResU-Net

ResU-Net is a variant of U-Net that brings together residual learning architectures and the encoder-decoder architecture. Deeper neural networks can lead to more improved performance; however, they can also increase model complexity thus making it difficult to train the model. He et al., [14] introduced the residual block to address this problem. The residual module helps in extracting deep features effectively and eliminating the vanishing gradient problem. A base U-Net architecture uses 3×3 convolutions with ReLu activation function in the encoder block, however, the ResU-Net replaces this with a pre-activated residual block[15]. The decoder block of the ResU-Net is similar to that of the standard U-Net. The residual block does not add overheads in the form of parameters to the network thus greatly improving the training speed and the model's efficiency[16].

2.4 TransU-Net

TransU-Net combines the strengths of Transformers and U-Net, making it a powerful alternative for medical image segmentation. The "Transformer encodes tokenized image patches derived from a CNN feature map to extract global context, while the decoder up samples the encoded features and merges them with high-resolution CNN feature maps, facilitating precise localization "[17]. Transformers originally introduced by Vasuwani et al., [18], have a self-attention mechanism which makes them ideal for sequence-to-sequence prediction. Combining transformers with U-Net address the U-Net challenge of limited ability to explicitly model long-range dependencies. According to [17], transformers are highly generalizable and therefore, ideal for segmentation models that may need to do downstream transfer learning. TransU-Net employs a CNN-Transformer hybrid model where CNN is first used as a feature extractor to generate a feature map for the input to the transformer[17].

2.5 V-Net

V-Net was proposed by Milletari et al., [19] as an approach that uses volumetric fully convolutional neural networks to perform 3D image segmentation. V-Nets were developed out of necessity. Standard U-Net could only perform 2D image segmentation, therefore, with the emergence of imaging machines that could generate 3D medical images, there was need to have a network that segment these images. Just like a standard U-Net the V-Net has an encoder and decoder paths. However, for the V-Net the convolutions performed in each stage use volumetric kernels having 5×5×5 voxels[19]. V-Nets although initially designed for 3D image segmentation they can be adapted to 2D image segmentation by modifying their 3D convolution blocks to 2D equivalents.

2.6 U-Net 3+

Introduced by Huang et al., [20], U-Net3+ was advancement of U-Net++ by providing full scale skip connections and deep supervisions. High level semantics and low-level details from feature maps constitute the full-scale skip connections. Deep supervision involves learning hierarchical representation from the full-scale aggregated feature maps. This architecture is suitable or objects such as organs that appear in varying scale[20]. The full-scale skip connections involve not only the interconnection between the encoder and decoder subnetworks but also the intra-connection within the decoder subnetwork. Huang et al., [20], further argue that the U-Net3+ reduces network parameters and improves computational efficiency.

3. Methodology

3.1 Dataset:

The study used two publicly available dataset that were obtained from Kaggle repository; The Breast Ultrasound dataset and the Brain MRI segmentation dataset. The Breast Ultrasound Images dataset was collected in 2018 and consists of breast ultrasound images of 600 female patients aged between 25 and 75 years old. The total number of images in the collection is 780 images grouped into normal (133 images), benign (487 images), and malignant (210 images) [21]. The Brain MRI segmentation dataset version available in Kaggle was originally obtained from The Cancer Imaging Archive (TCA). It contains brain MRI images together with their Fluid-Attenuated Inversion Recovery (FLAIR) abnormality segmentation masks. FLAIR is a specific MRI (Magnetic Resonance Imaging) modality commonly used in medical imaging, particularly for brain scans. This technique is designed to suppress the signals from fluids like cerebrospinal fluid (CSF), making it easier to detect abnormalities such as lesions, tumors, edema, and demyelination. The dataset corresponds to 110 patients and has been previously used in other studies such as[22].

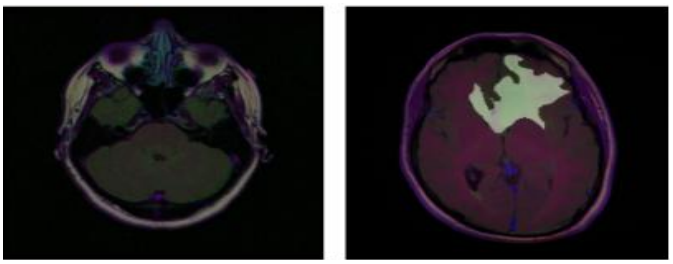


Figure 1: Sample Brain MRI Images

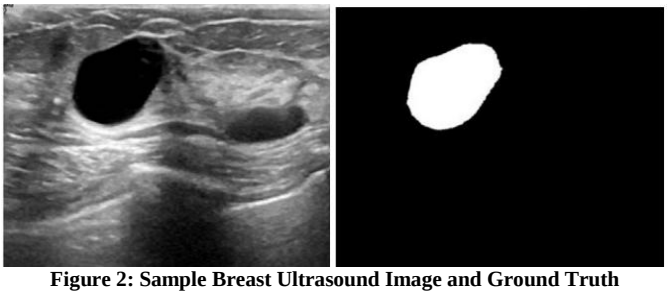


Figure 2: Sample Breast Ultrasound Image and Ground Truth

3.2 Data Pre-processing

Data pre-processing involved resizing the images to size 256*256*1 for breast ultrasound dataset and size 256*256*3 for the Brain MRI datasets. This was followed by mapping the images to their respective masks, then data augmentation was done to prevent the models from overfitting. The augmentation done to the dataset include, rotation, zoom, height shift, width shift, and horizontal flip. A train-validation-test split at the ratio of 0.8:0.1:0.1 respectively was

© 2025, IJSRCSE All Rights Reserved

applied to both datasets.

3.3 Models' Architectures

In this comparative study, the researcher standardized the model architectures and training parameters across all U-Net variants for each dataset. For the Breast Ultrasound dataset, all networks featured 4 convolutional blocks in the encoder path, a bottleneck, and 4 blocks in the decoder path. The output layer consisted of a Conv2D that reduced the feature maps to a single channel with a sigmoid activation. The models were trained using the cross-entropy loss function, Adam optimizer with a default learning rate of 0.001, for 100 training epochs. Similarly, in the Brain MRI segmentation dataset, the networks comprised an encoder with four blocks, a bottleneck, and a decoder with four blocks. These models were trained using the dice loss function and the Adamax optimizer with a learning rate of 0.001 for 100 epochs. The models were developed using TensorFlow framework with Keras in the backend and trained using Kaggle notebooks with GPU P100.

3.4 Model Evaluation

Three evaluation metrics namely; Accuracy, Jaccard index also known as Intersection over Union, and Dice Similarity coefficient were used.

Accuracy refers to the number of correct predictions which can be computed as shown below

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

Where TP =True Positive, TN= True Negative, FP= False Positive, and FN is the False Negative.

Dice coefficient measures the similarity between the predicted and true regions, focusing on overlap. It can be computed as follows:

$$Dice = \frac{2TP}{2TP + FP + FN} \quad (2)$$

Jaccard index is the area of the intersection over union of the predicted segmentation and the ground truth. It can be computed as follows:

$$Jaccard\ (IoU) = \frac{Area\ of\ Overlap}{Area\ of\ Union} \quad (3)$$

4. Results and Discussion

The results of the experiments are as shown in table 1. The results indicate that ResU-Net achieved the highest performance in all metrics in the breast Ultrasound image dataset. The ResU-Net model also emerged as the best in Jaccard index and Dice Similarity index in the Brain MRI segmentation dataset. U-Net++ achieved the highest accuracy of 99.72% in the Brain MRI Segmentation dataset. The best performance per metric in each dataset is highlighted in bold.

Table 1: Performance Results on the Breast Ultrasound Dataset

Architecture	Accuracy (%)	Jaccard Index	Dice Similarity Index
ResU-Net	**96.20**	**0.5158**	**0.5930**
TransU-Net	95.09	0.4048	0.5099
U-Net++	95.82	0.4886	0.5771
U-Net	92.76	0.1060	0.1805
U-Net3+	92.20	0.0490	0.0912
V-Net	95.78	0.4793	0.5684

Table 2: Performance Results on the Brain MRI Segmentation Dataset

Architecture	Accuracy (%)	Jaccard Index	Dice Similarity Index
ResU-Net	99.70	**0.7923**	**0.8730**
TransU-Net	99.01	0.0973	0.0973
U-Net++	**99.72**	0.7882	0.8694
U-Net	99.53	0.7905	0.6751
U-Net3+	98.99	0.0655	0.0655
V-Net	99.70	0.7348	0.8214

The loss curves of the two top-performing models on the Brain MRI dataset as shown in Figure 3 reveal that, although both models' training and validation losses converged after epoch 90, ResU-Net exhibited inconsistent training loss in earlier epochs. This suggests that the common learning rate used was not optimal for ResU-Net between epochs 45 and 85. Future research could explore using a dynamic learning rate that adjusts based on loss reduction, which might improve stability during training. This inconsistency could explain why ResU-Net ranked second in accuracy on this dataset.

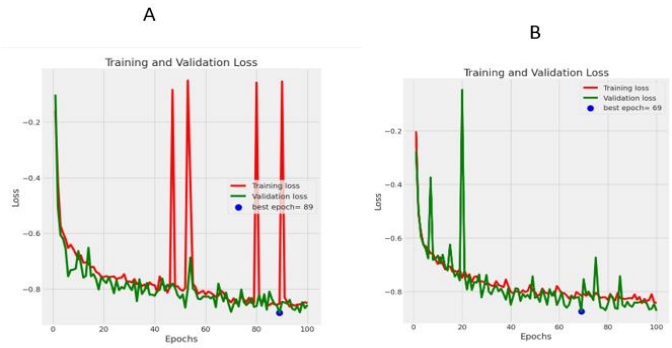


Figure 3: ResU-Net (A) and U-Net++ (B) Training and Validation Loss in the Brain MRI segmentation dataset

The results indicate that U-Net variants are highly effective convolutional neural networks for medical image segmentation. Despite architectural differences, each variant exhibited strong performance, with only slight variations in their results. The variance is mainly be due to some architectural nuances in the models. Although the Artificial neural theory[23], [24] forms the bedrock of designing neural networks, the information theory is the one that deals how

© 2025, IJSRCSE All Rights Reserved

neural networks learn[25]. Different architectures have different capabilities of gaining maximum of entropy of information as per the information theory[25]. Therefore, the better the architecture is at extracting and propagating suitable features the higher the models' performance.

The superior performance of ResU-Net can be largely attributed to its use of residual blocks, which helps to mitigate the vanishing gradient problem[14]. Residual blocks also serve as good feature extractors which is ideal for semantic segmentation since they not only extract foreground and background but also segment each pixel in the image. Residual networks have skip connections that act as meta-learner by bypassing some layers during training thus enabling the model to generalize quickly[26]. This aligns with the curriculum learning theory[27] since it provides the ResU-Net model with a learning strategy that starts with less complex examples and gradually increases the complexity of examples thus increasing information gain and minimizing information entropy.

U-Net++ enhances multi-scale feature fusion and optimizes information flow thus minimizing information entropy. The nested skip pathways refine low-level features before merging them with high-level features. This enhances mutual information between different scales, leading to better spatial localization of segmented regions. This explains why overall it was the first runners-up in terms of performance.

U-Net3+ showed the lowest performance across all metrics on both datasets, likely due to the high complexity of intra-connections in its decoder block, highlighting that increased architectural complexity does not always translate to better performance. The Intra-connections may require a better training strategy in order to align with provision of curriculum learning, therefore, the universal learning approach used here did not minimize the information entropy of Unet3+.

V-Net's accuracy matched that of ResU-Net in the Brain MRI segmentation dataset, demonstrating that V-Net, originally designed for 3D images, can also perform well on 2D data. TransU-Net did not achieve high Jaccard Index and Dice Coefficient scores in the Brain MRI segmentation task, likely due to the increased complexity introduced by the encoding of CNN features into the transformer block. The self-attention mechanism[28] used in transformers require much bigger dataset and a more robust data processing pipeline for vision transformers to deliver good performance.

5. Conclusion and Future Scope

This study presents an empirical analysis of six variants of the encoder-decoder U-Net architecture to evaluate their performance in image segmentation under a consistent experimental setup and different imaging modalities. The results demonstrate that all six architectures are highly effective, with ResU-Net showing the best overall performance, closely followed by U-Net++. The superior performance of ResU-Net is largely attributed to the advantages of its residual block design, which helps address challenges like the vanishing gradient problem. This study offers a valuable reference point for researchers in medical image processing, enabling them to select the most suitable architecture based on their specific dataset requirements. Future research can explore the impact of adaptive learning rates, alternative loss functions, and the integration of more robust learning strategies to further enhance the performance of these U-Net variants. Additionally, investigating the application of these architectures on larger and more diverse medical datasets, including 3D data, could provide deeper insights into their robustness and generalizability. Researchers can also explore how intelligent agent theory [29] can be applied in medical image processing as a path towards artificial general intelligence.

Data Availability
The datasets used in this study can be accessed from Kaggle repository

Conflict of Interest
The authors have no conflict of interest to declare

Funding source
None

Authors contribution
All authors contributed to the experimentation, manuscript editing, revision and have approved the final version of the manuscript.

Acknowledgement
None

References

[1] R. Azad *et al.*, "Medical Image Segmentation Review: The Success of U-Net," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1–20, 2024, doi: 10.1109/TPAMI.2024.3435571.

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

[3] J. Kamiri, G. M. Wambugu, and A. M. Oirere, "Multi-task Deep Learning in Medical Image Processing: A Systematic Review," *Int. J. Comput. Sci. Res.*, vol. 9, pp. 3418–3440, Jan. 2025.

[4] R. Azad *et al.*, "Medical Image Segmentation Review: The Success of U-Net," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1–20, 2024, doi: 10.1109/TPAMI.2024.3435571.

[5] T. Zhou, S. Ruan, and S. Canu, "A review: Deep learning for medical image segmentation using multi-modality fusion," *Array*, vol. 3–4, p. 100004, Sep. 2019, doi: 10.1016/j.array.2019.100004.

[6] Z. Guo, X. Li, H. Huang, N. Guo, and Q. Li, "Deep Learning-Based Image Segmentation on Multimodal Medical Imaging," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 3, no. 2, pp. 162–169, Mar. 2019, doi: 10.1109/TRPMS.2018.2890359.

[7] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," Springer International Publishing, 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.

[8] W. Ehab and Y. Li, "Performance Analysis of UNet and Variants for Medical Image Segmentation," arXiv.org.

[9] A. K. Sahu, P. K. Gupta, A. K. Singh, K. Singh, and U. Kumar, "A Convolutional Neural Network Approach for Detecting the Distracted Drivers," *Int. J. Sci. Res. Comput. Sci. Eng.*, vol. 11, no.

© 2025, IJSRCSE All Rights Reserved

1, pp. 1–6, Feb. 2023.
[10] 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.
[11] A. H. Alharbi, C. V. Aravinda, M. Lin, P. S. Venugopala, P. Reddicherla, and M. A. Shah, "Segmentation and Classification of White Blood Cells Using the UNet," *Contrast Media Mol. Imaging*, vol. 2022, no. 1, p. 5913905, 2022, doi: 10.1155/2022/5913905.
[12] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A Nested U-Net Architecture for Medical Image Segmentation, Springer International Publishing, 2018, pp. 3–11. doi: 10.1007/978-3-030-00889-5_1.
[13] N. Micallef, D. Seychell, and C. J. Bajada, "Exploring the U-Net++ Model for Automatic Brain Tumor Segmentation," *IEEE Access*, vol. 9, pp. 125523–125539, 2021, doi: 10.1109/ACCESS.2021.3111131.
[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA: IEEE, Jun. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.
[15] M. W. Sabir *et al.*, "Segmentation of Liver Tumor in CT Scan Using ResU-Net," *Appl. Sci.*, vol. 12, no. 17, Art. no. 17, Jan. 2022, doi: 10.3390/app12178650.
[16] C. Chen, L. Jing, H. Li, and Y. Tang, "A New Individual Tree Species Classification Method Based on the ResU-Net Model," *Forests*, vol. 12, no. 9, Art. no. 9, Sep. 2021, doi: 10.3390/f12091202.
[17] J. Chen *et al.*, "TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation," Feb. 08, 2021, *arXiv*: arXiv:2102.04306. doi: 10.48550/arXiv.2102.04306.
[18] A. Vaswani *et al.*, "Attention is All you Need," in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2017.
[19] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," Jun. 15, 2016, *arXiv*: arXiv:1606.04797. doi: 10.48550/arXiv.1606.04797.
[20] H. Huang *et al.*, "UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation," Apr. 19, 2020, *arXiv*: arXiv:2004.08790. doi: 10.48550/arXiv.2004.08790.
[21] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, "Dataset of breast ultrasound images," *Data Brief*, vol. 28, p. 104863, Feb. 2020, doi: 10.1016/j.dib.2019.104863.
[22] M. Buda, A. Saha, and M. A. Mazurowski, "Association of genomic subtypes of lower-grade gliomas with shape features automatically extracted by a deep learning algorithm," *Comput. Biol. Med.*, vol. 109, pp. 218–225, Jun. 2019, doi: 10.1016/j.compbiomed.2019.05.002.
[23] O. A. Montesinos López, A. Montesinos López, and J. Crossa, "Fundamentals of Artificial Neural Networks and Deep Learning," Springer International Publishing, 2022, pp. 379–425. doi: 10.1007/978-3-030-89010-0_10.
[24] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.
[25] F. Radicchi, D. Krioukov, H. Hartle, and G. Bianconi, "Classical information theory of networks," *J. Phys. Complex.*, vol. 1, no. 2, p. 025001, Jul. 2020, doi: 10.1088/2632-072X/ab9447.
[26] J. Kamiri, G. M. Wambugu, and A. Oirere, "A Systematic Literature Review of Meta-Learning Models for Classification Tasks," 2022, doi: 10.7753/IJCATR1103.1002.
[27] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in *Proceedings of the 26th Annual International Conference on Machine Learning*, Montreal Quebec Canada: ACM, Jun. 2009, pp. 41–48. doi: 10.1145/1553374.1553380.
[28] A. Vaswani *et al.*, "Attention is All you Need," in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2017. Accessed: Jan. 23, 2024.
[29] S. K. Wanjau and S. N. Thiiru, "A Reinforcement Learning-Based Multi-Agent System for Advanced Network Attack Prediction," *Int. J. Sci. Res. Comput. Sci. Eng.*, vol. 12, no. 6, pp. 14–24, Dec. 2024.

AUTHORS PROFILE



Jackson Kamiri received a Bachelor of Business Information Technology (2019) and Master of Science in Information Technology (2022), from Murang'a University of Technology, Kenya. Currently, he is pursuing Ph.D. Information Technology and serving as a tutorial fellow in the department of Information Technology, Murang'a University of Technology. His research interests include deep learning, Image processing, and Big data analytics.



Geoffrey Mariga Wambugu received his B.Sc. in Mathematics and Computer Science (2000) from Jomo Kenyatta University of Agriculture and Technology (JKUAT), Kenya. An M.Sc. in Information Systems (2019) from The University of Nairobi, Kenya, and the Ph.D. in Information Technology (2019) from JKUAT. Currently, he is a Senior Lecturer in the Information Technology department at Murang'a University of Technology. His research interests include Probabilistic Machine Learning, Natural Language Processing, Data mining, Deep learning, Big Data Analytics.



Aaron Mogeni Oirere received his Bsc Computer Science from Periyar University, India, in 2007, and M.Sc. Computer Science from Bharathier University, India, in 2010 and Ph.D. Computer Science from Dr. Babasaheb Ambedkar Marathwada University, India. He is currently a lecturer and the COD of the computer Science department at Murang'a University of Technology. His research interests include Deep learning, Text analytics, Speech recognition, Human computer Interaction, and Image processing.

© 2025, IJSRCSE All Rights Reserved