



Evaluating the Impact of Hyperparameters Using U-Net Networks with Transfer-Learning for Cerebellum Segmentation from Ultrasound Scans

Amisha Sakariya, Jayesh Chaudhary, Bhavesh Patel, Jaydeep Barad

Department of Computer Engineering, Sarvajani College of Engineering and Technology, Gujarat, India

E-mails: amishasakariya.comtech23@scet.ac.in

jayesh.chaudhary@scet.ac.in

bhavesh.patel@scet.ac.in jaydeep.barad@scet.ac.in

Abstract: *The cerebellum is an essential part of the human brain, responsible for motor learning, balance, and coordination. Therefore, cerebellar growth in the fetal brain is identified by its shape on ultrasound images. To overcome the limitations of the current clinical process, several computerized approaches have been proposed in the literature. This research paper proposes cerebellum segmentation from 2D ultrasound images of the fetal brain using U-Net, Attention U-Net, and transfer-learning-based U-Net models, namely VGG16-U-Net, ResNet50-U-Net, and EfficientNetB1-U-Net, with (1) four different evaluation parameters: Dice coefficient, Intersection over Union (IoU), Precision, and loss, (2) three different loss functions, namely Binary Cross-Entropy (BCE) loss, Dice loss, and Tversky loss, (3) three batch sizes – 8, 16, and 32. Lastly, the effectiveness of the trained models for cerebellum segmentation is assessed for real-time data. This research aims to segment the cerebellum in second-trimester fetal brain ultrasound images to predict prenatal neurodevelopmental disorders.*

Keywords: *Attention U-Net, Fetal cerebellum, Segmentation, Transfer learning, U-Net.*

1. Introduction

One commonly used technique to track fetal growth and development is ultrasound imaging. Ultrasound imaging provides several advantages, including high reliability, low cost, portability, and real-time visualization. The cerebellum, cerebrum, midbrain, and thalamus are among the fetal brain regions measured on an

ultrasound scan as part of the fetal anomaly screening process, which is carried out during the second trimester of pregnancy [1, 5, 8]. The cerebellum, which means “little brain” in Latin, is popular for its role in motor control [2, 35]. The cerebellum is situated above the brainstem and under the cerebrum at the base of the brain. Although occupying only 10% of the brain’s volume, the cerebellum is linked to almost half of the body’s neurons. The ultrasound modality maintains integrity throughout fetal development and can differentiate surrounding brain structures. Additionally, ultrasound is a non-invasive diagnosis technique and will give real-time sonograms of the internal organs and tissues with no harmful radiation for the fetus. Therefore, during the anomaly scan, the cerebellum is examined using ultrasound. The cerebellum is a major structure for understanding neurodevelopmental consequences and identifying prenatal abnormalities [12]. Cerebellar abnormalities include cerebellar hypoplasia, cisterna magna, Dandy-Walker syndrome, and COACH syndrome, which may occur during fetal brain development [6-8, 22]. Therefore, evaluation of cerebellar growth during development is crucial.

Ultrasound images may exhibit a variety of imperfections, including shadows, speckle noise, motion blurring, and missing borders, due to complex interactions between ultrasonic waves and the patient’s biological tissues and the fetus [33, 34]. Therefore, analyzing fetal ultrasound images is challenging. Examining the boundaries of the cerebellum from the ultrasound images requires more time and considerable clinical expertise and experience. In many cases, it leads to variability between and within observers and depends on the sonographer’s experience [13, 25]. Additionally, if the fetal position is not correct during the anomaly scan, it can be challenging for sonographers to obtain a clear image of the desired view. The patient’s body fat and fetal position play a major role in the quality of ultrasound images. These factors can result in low signal-to-noise ratios, acoustic shadows, or artifacts, which, if misread, could lead to diagnostic errors [14]. There is still a shortage of skilled obstetricians, despite recent improvements in medical care. For this purpose, a computerized approach for segmenting the cerebellum in the fetal brain is necessary to determine its shape without the subjectivity of different sonographers. Thus, our objective is to develop a computerized approach for semantic segmentation of the cerebellum using U-Net variants and to apply transfer learning to the contracting path of the U-Net.

2. Literature review

Less work has been done on the segmentation of fetal brain cerebellar ultrasound images. However, segmentation of the human cerebellum from Magnetic Resonance Imaging data has been completed. Additionally, research has been conducted on other parts of the brain using ultrasound images. They propose machine-learning-based, CNN-based, or morphological operations [16]. Machine learning approaches include the watershed algorithm [9], Artificial neural networks [10], Support vector machines [10], the Hough Transform [17], Dynamic Programming [17], and Ellipse Fit [17]. Machine learning-based segmentation approaches require significant time, which is a major drawback. Additionally, the watershed algorithm requires careful adjustment to select the right thresholds.

Numerous CNN-based approaches have been proposed in [11, 15, 18, 24, 25, 36] to address the drawbacks of traditional segmentation approaches. The proposed architecture, ResU-Net-c [11], combines features from both U-Net and ResNet for cerebellar segmentation in 2D fetal ultrasound images. The inclusion of residual blocks enhances speed and enables the development of deeper networks capable of reconstructing the cerebellum with clear, precise boundaries. It uses long skip connections rather than short skip connections within the block. Also, uses a dilated convolution operation to retain the full spatial dimensions.

Lu et al. [36] proposed ECAU-Net, a modified version of the traditional U-Net that integrates an ECA module into its architecture. Coronado González et al. [18] came up with a multi-step technique for preprocessing and cerebellum segmentation from MRI. It includes the extraction of the cerebellum mask, reducing the variability of irregular intensity using N4 Bias Field Correction [20], and the MRI's spatial normalization by registering it to the ICBM 2009c template in MNI space [21]. Lee et al. [23] proposed a cerebellum segmentation approach for brain MRI that combines binary brain masks and Partial-Volume Estimation (PVE) maps. Morphological operations and mask-refinement techniques are used to segment precise, accurate cerebellar boundaries. Walsh et al. [24] proposed a lightweight U-Net architecture for segmenting brain tumors from MRI images, using a smaller kernel size and limited depth. This research shows that without a large dataset or data augmentation, an optimized U-Net can accurately segment brain tumors. However, the model's generalization may be affected by the loss of some 3D contextual information in 2D slices.

Ali et al. [25] demonstrated an ensemble deep learning method for segmenting brain tumors in multimodal MRI images, which combines a 3D CNN with a 2D U-Net architecture. Jang et al. [15] presented a computerized approach

to determine fetal growth and gestational age by automatically assessing the fetal Abdominal Circumference (AC) biometric indicator. It classifies five anatomical structures: the Stomach Bubble (SB), the Umbilical Vein (UV), the Amniotic Fluid (AF), and the Shadowing Artifact (SA). Then, the results are used to locate important anatomical structures, followed by the Hough Transform to fit ellipses and estimate the fetal Abdominal Circumference (AC).

Thomas, Peters and Jeanty [16] proposed an approach to automatically measure fetal femoral length. Morphological filtering techniques are applied to the input ultrasound images to produce a progressively varying background image. Then the background of the original image is removed, resulting in a clearer, higher-contrast image. It uses morphological opening and closing operators to separate the femur from the image. The weight of each detected region is calculated after these morphological operations, and the region with the highest weight is selected as the femur area. The femur area is then simplified to its skeletal form using a hit-or-miss morphological operator. The skeleton contains two endpoints of the femur bone; thus, the femur length is produced by dividing the pixel count between the two endpoints of the femur bone by pixel length to millimeters.

Van den Heuvel et al. [17] proposed an approach to estimate fetal Head Circumference (HC) from ultrasound images acquired across all three trimesters of pregnancy. It consists of a two-step process. The initial step involves locating the skull using a random forest classifier trained on Haar-like features. The next step estimates head circumference by selecting the best result from the Hough transformation, dynamic programming, and the eclipse fit method. The main limitation of this research is that it is based on data collected from a single hospital.

The segmentation techniques mentioned above have a number of limitations: 1) Transfer learning encoders have not been used to test any of them; 2) A single loss function, binary cross-entropy loss, has been used to train the models; 3) the models have not been experimented with different batch sizes and optimizers; 4) A smaller dataset has been used in some of the approaches to train the model.

According to the literature, classic U-Net models often require many iterations, making them time-consuming. For accurate, fast segmentation, efficient hyperparameter tuning is required. Researchers have begun incorporating transfer learning into the U-Net framework to address these problems. By using weights from pre-trained CNNs in the encoder block of the U-Net architecture, these hybrid models improve the segmentation model's overall performance by lowering the number of trainable parameters and training iterations. A redesigned U-Net architecture [27] has just been shown to enhance feature extraction. The region of

interest has been identified in [28] using a hybrid architecture that combines U-Net and VGG16.

Among the previously discussed segmentation techniques, the cerebellum in fetal brain ultrasound images has been segmented using the methods described in [11] and [36]. The cerebellum has been segmented from Magnetic Resonance Imaging (MRI) using several techniques [18, 23]. Brain tumors from MRI have been segmented using additional techniques proposed in [24, 25]. Additionally, other areas of the fetal brain have been segmented from ultrasound images using the techniques described in [15-17]. Thus, this study demonstrates cerebellar segmentation of the fetal brain from ultrasound images using U-Net and its variants.

3. Cerebellum segmentation using U-based architectures

The U-Net and its variants, which we employ in our study to segment the cerebellum from 2D ultrasound images, are discussed in this section.

By splitting the image into larger regions, segmentation removes the need to process each pixel independently. There are two common segmentation techniques: semantic segmentation and instance segmentation. Pixels are arranged using semantic segmentation based on similar categories or interpretations. On the other hand, instance segmentation treats even similar objects as distinct entities, identifying and tracking each pixel as belonging to a unique individual.

Medical image segmentation commonly uses encoder-decoder architectures such as U-Net and Attention-U-Net. The encoder consists of several convolutional layers that produce a feature map by downsampling the image's spatial resolution. The segmentation results are obtained by upsampling the feature map using the decoder, which includes transpose convolutional layers. Various improved versions of U-Net have been proposed by applying transfer learning, using pre-trained weights from well-known CNN architectures such as VGG16, ResNet50, and EfficientNetB1, and incorporating them into the encoder to enhance performance.

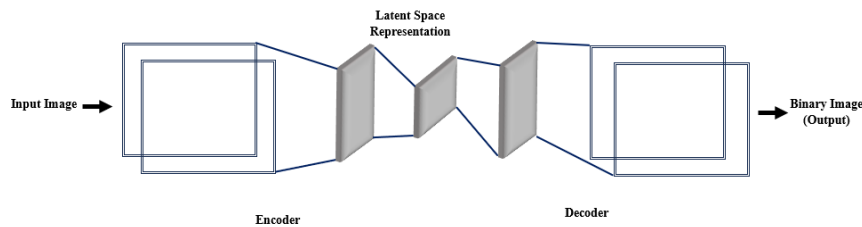


Fig. 1. CNN for image segmentation using an encoder-decoder architecture

3.1. U-Net Architecture

The traditional CNN architecture, U-Net, serves as the foundation for the proposed structure [19], which leverages earlier feature maps. This feature allows the model to handle more complex images while enabling the network to do the same. The U-Net architecture consists of two main parts: a contracting path, also known as the encoder, and an expanding path, also known as the decoder. Additionally, a contracting path and an expanding path are connected through a bottleneck, forming a symmetrical U-shaped architecture. The U-Net architecture with both parts is shown in Fig. 2, and its architectural components are illustrated in Table 1. The main characteristic of the U-net architecture is the integration of skip connections. Their main task is to transmit the spatial information from the contractive path to the corresponding expanding path. Algorithm 1 describes the encoder-decoder process used to segment the cerebellum using the U-Net architecture, and Table 2 describes the phrases used in the algorithm.

Algorithm 1. U-Net-Based Cerebellum Segmentation Approach Pseudocode

Input: A set U of Ultrasound images of the Cerebellum of the Fetal Brain of N patients of the second trimester $U = \{U_1, U_2, \dots, U_N\}$

Output: Segmented Cerebellum Mask image of the Fetal Brain

BEGIN

 // Contractive path of U-Net

FOR $i = 1$ to n layers

$U'_i = \text{Conv2D}(U_{i-1}, f_i, k_i, s_i, p_i)$

$U'_i = \text{BatchNormalization}(U'_i)$

$U'_i = \text{ReLU}(U'_i)$ // Apply activation function

$T_i = U'_i$ // Store U'_i in T_i to further use in expansive path

$U'_i = \text{MaxPooling2D}(U'_i, k_p, s_p)$

END FOR

 // Bottleneck

$U_{\text{bottleneck}}' = \text{Conv2D}(U_n, f_{n+1}, k_{n+1}, s_{n+1}, p_{n+1})$

$U_{\text{bottleneck}}' = \text{BatchNormalization}(U_{\text{bottleneck}}')$

$U_n = \text{ReLU}(U_{\text{bottleneck}}')$

 // Expansive path of U-Net

FOR $j = n$ to 1

$U'_j = \text{Conv2DTranspose}(U_{j+1}, f_j, k_t, s_t, p_t)$

$U'_j = \text{Concatenate}(U'_j, T_j)$

$U'_j = \text{Conv2D}(U'_j, f_j, k_j, s_j, p_j)$

$U'_j = \text{BatchNormalization}(U'_j)$

$U_j' = \text{ReLU}(U_j') \text{ // Apply activation function}$
 END FOR
 $O = \text{Conv2D}(U_j', f_{\text{output}}, k_{\text{output}}, s_{\text{output}}, p_{\text{output}}) \text{ // Last convolutional layer}$
 $O_{\text{output}} = \text{sigmoid}(O) \text{ // Apply activation function}$
 END

Table 1. Details of U-Net architecture

Level	Layer	Number of filters	Input size	Output size
Contracting path	Input layer	-	(256, 256, 1)	(256, 256, 1)
	Layer 1 (Conv2D + BN + ReLu + Dropout + Conv2D + BN + ReLu)	16	(256, 256, 1)	(256, 256, 16)
	MaxPooling2D	-	(256, 256, 16)	(128, 128, 16)
	Layer 2 (Conv2D + BN + ReLu + Dropout + Conv2D + BN + ReLu)	32	(128, 128, 16)	(128, 128, 32)
	MaxPooling2D	-	(128, 128, 32)	(64, 64, 32)
	Layer 3 (Conv2D + BN + ReLu + Dropout + Conv2D + BN + ReLu)	64	(64, 64, 32)	(64, 64, 64)
	MaxPooling2D	-	(64, 64, 64)	(32, 32, 64)
	Layer 4 (Conv2D + BN + ReLu + Dropout + Conv2D + BN + ReLu)	128	(32, 32, 64)	(32, 32, 128)
	MaxPooling2D	-	(32, 32, 128)	(16, 16, 128)
	Layer 5 (Conv2D + BN + ReLu + Dropout + Conv2D + BN + ReLu)	256	(16, 16, 128)	(16, 16, 256)
Expansive path	ConvTranspose2D	128	(16, 16, 256)	(32, 32, 128)
	Concatenate	-	(32, 32, 128) + Layer 4	(32, 32, 256)
	Layer 6 (Conv2D + BN + ReLu + Dropout + Conv2D + BN + ReLu)	128	(32, 32, 256)	(32, 32, 128)
	ConvTranspose2D	64	(32, 32, 128)	(64, 64, 64)
	Concatenate	-	(64, 64, 64) + Layer 3	(64, 64, 128)
	Layer 7 (Conv2D + BN + ReLu + Dropout + Conv2D + BN + ReLu)	64	(64, 64, 128)	(64, 64, 64)
	ConvTranspose2D	32	(64, 64, 64)	(128, 128, 32)
	Concatenate	-	(128, 128, 32) + Layer 2	(128, 128, 64)
	Layer 8 (Conv2D + BN + ReLu + Dropout + Conv2D + BN + ReLu)	32	(128, 128, 64)	(128, 128, 32)
	ConvTranspose2D	16	(128, 128, 32)	(256, 256, 16)
	Concatenate	-	(256, 256, 16) + Layer 1	(256, 256, 32)
	Layer 9 (Conv2D + BN + ReLu + Dropout + Conv2D + BN + ReLu)	16	(256, 256, 32)	(256, 256, 16)
	Output layer (Conv2D + sigmoid)	1	(256, 256, 16)	(256, 256, 1)

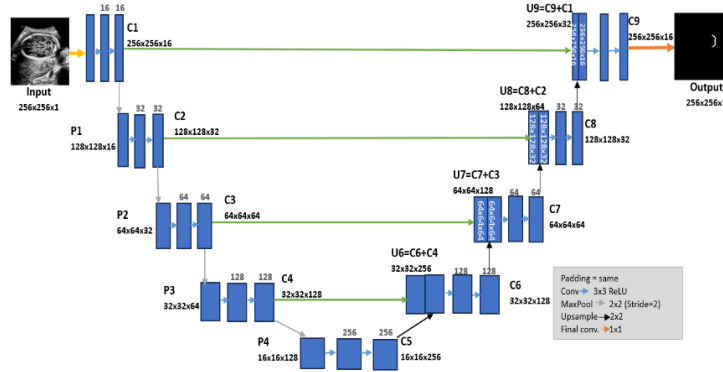


Fig. 2. U-Net architecture for image segmentation

Table 2. Notation list

Symbol	Notation
U	A set of ultrasound images
N	Total number of images
n	Number of total layers (depth of U-Net)
f_i	Number of total filters at layer i
k_i	Kernel size at layer i
s_i	Size of stride at layer i
p_i	Padding at layer i
T_i	Temporary for storing U_i'
k_p	Pooling kernel size
s_p	Size of Pooling stride
k_t	Transpose convolution kernel size
s_t	Transpose convolution stride
p_t	Transpose convolution padding
O	Output of the last convolution layer
f_{output}	Number of output filters
k_{output}	Kernel size at the output layer
s_{output}	Stride at the output layer
p_{output}	Padding at the output layer
O_{output}	Segmented output image

3.2. Attention-U-Net architecture

To improve segmentation performance, particularly for complex medical images, Attention U-Net, an extension of the classic U-Net design [29-31], integrates attention mechanisms into the skip connections. Attention U-Net includes attention gates (AGs) [30, 31] that selectively highlight relevant spatial regions [26] and suppress irrelevant background features before feature fusion. In contrast, traditional U-Net concatenates encoder and decoder feature maps directly.

The attention gates use contextual information from the decoder to influence the encoder features, allowing the network to focus on essential structures [30]. In most cases, target areas such as organs, lesions, or tumors occupy only a small fraction of the image and are usually surrounded by noisy or low-contrast

backgrounds. Attention mechanisms have been proven to be advantageous [29]. Attention U-Net increases localization accuracy and boundary segmentation by fine-tuning feature results without significantly increasing computing complexity [29, 31].

In today's scenario, the attention U-Net has surpassed the traditional U-Net [30] for medical images, especially for small, irregular, or low-contrast anatomical regions. By eliminating extraneous activations, it reduces false positives and improves performance, while accounting for the most common problem in clinical practice: a limited annotated dataset.

3.3. U-Net Architectures with transfer-learning

The traditional U-Net and the Attention U-Net are well-known for their usefulness in medical image segmentation. Due to their encoder-decoder structure, they can preserve spatial details. However, this increases computational complexity and training time, as U-Net models trained from scratch entail many trainable parameters [29]. To overcome these problems, transfer learning is used to improve the coherence rate and segmentation accuracy.

Transfer learning is the process of retraining a pre-trained deep learning model on a target dataset after it has been trained on a big dataset to capture both generic and specialized image attributes. These pre-trained models are then adjusted for specific segmentation tasks. There are two main advantages to using transfer learning for segmentation [3]. First of all, it performs better than building models from scratch, particularly when the dataset size used is smaller. Second, because the model already has well-developed feature representations, it significantly lowers training time and computing cost [32].

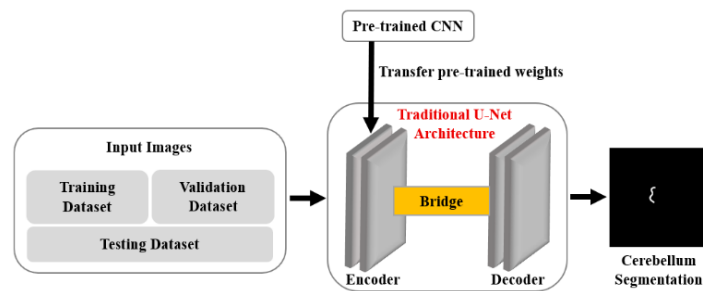


Fig. 3. Diagrammatic illustration of the U-Net approaches based on transfer learning

In this research, U-Net is trained using transfer learning with VGG16, ResNet50, and EfficientNet-B1 as encoders for medical image segmentation. The encoder path of the U-Net architecture integrates pre-trained CNN models, while the decoder is trained to generate high-resolution segmentation maps that align with

the desired regions. The schematic representation of the transfer-learning-based U-Net approach is shown in Fig. 3.

4. Experiments and results

This section describes the implementation tools, dataset, experimental setup, loss functions, optimizers, batch sizes, evaluation metrics, and the results and their analysis.

4.1. Dataset

There is no publicly available dataset of ultrasound images of the fetal cerebellum, as research in this domain is limited. We obtained a dataset acquired from a Sonoscape P40 Elite ultrasound machine in Gujarat, India, under the supervision of an experienced, qualified radiologist. We obtained a dataset of 328 2D ultrasound images of the fetal cerebellum, mainly captured during the fetal anomaly scan in the second trimester of pregnancy. The ultrasound images are 880×1264 pixels in dimension. Additionally, all private patient information has been removed from the ultrasound images. All images are in RGB color space. Semantic labels are labeled by an expert radiologist with the open-source software Label Studio. Table 3 describes the dataset. However, Fig. 4 shows ultrasound images of the fetal brain cerebellum.

Table 3. Dataset description

Symbol	Notation
Type of data	Ultrasound images
Anatomy category	Cerebellum of the fetal brain
Source of the dataset	Krishna X-ray and sonography clinic
Image source	Sonoscape P40 Elite US machine
No. of images	328
Image dimensions	880×1264
Image extension	.jpg
Image colorspace type	RGB images

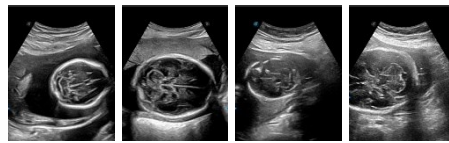


Fig. 4. Samples of the ultrasound images of the fetal brain cerebellum

4.2. Experimental setup

The experiments are conducted in three phases, as illustrated in Fig. 5: data preparation, model training, and model deployment. The collected ultrasound images of the fetal brain's cerebellum are processed in the first phase.

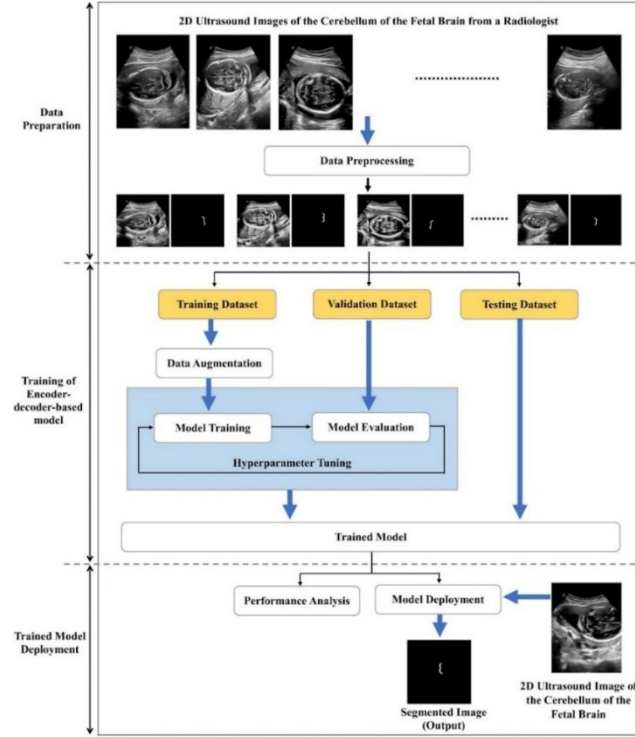


Fig. 5. Proposed segmentation pipeline using U-Net and its versions

The data preparation phase consists of four steps designed to enhance image quality and accelerate model training. The images are initially resized to a uniform resolution to standardize the input dimensions and reduce the computational cost of the training. Subsequently, the images are converted to grayscale, reducing the number of channels and, in turn, decreasing computational cost while preserving crucial structural and anatomical details. To enhance image quality, median filtering is used to reduce speckle noise, a frequent artifact in ultrasound imaging, without losing edge details. Finally, Contrast-Limited Adaptive Histogram Equalization (CLAHE) is used to enhance local contrast and highlight finer structural information, thereby improving feature visibility for the learning process. After preprocessing the images, corresponding binary mask images are created under the guidance of an experienced sonographer. In order to enable faster and more efficient

training, the pixel values of the input images are normalized to the range of 0 to 1 before feeding the data to the model.

During the second phase, the model is trained using pre-processed images. 70% (229) of the images are used for training the model, 20% (67) of the images are used for validation of the model, and 10% (32) of the images are used to test the trained model out of 328 ultrasound images. To generalize the model, we also apply augmentation techniques to the training data. To determine the optimal batch size that strikes the best trade-off between gradient stability and computational efficiency during training, we use batch sizes of 8, 16, and 32. The Rectified Linear Unit (ReLU) activation function is used in the hidden layer to improve the model's capacity to learn and represent features effectively, while incorporating nonlinearity. In the output layer, the sigmoid activation function is used to enable pixel-wise classification in segmentation by converting the hidden-layer predictions into probabilities in the range $[0, 1]$. To achieve stable, effective convergence during training, we use the Adam and RMSprop optimizers with a learning rate of 0.0001 for our dataset. We set the training epoch limit to 75 to avoid unnecessary computation. Additionally, we also set early stopping as a regularization approach.

The trained model is evaluated on the test dataset in the third phase to assess its performance on new and unseen data for segmenting the cerebellum.



Fig. 6. Training diagram of the proposed cerebellum segmentation approach

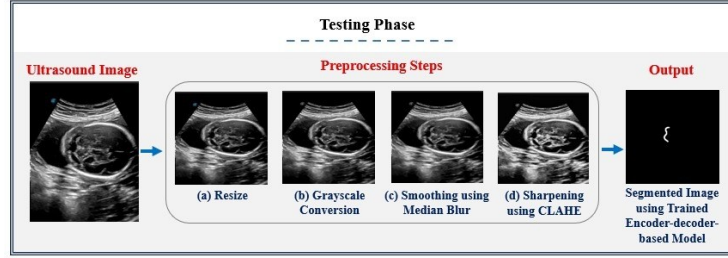


Fig. 7. Testing diagram of the proposed cerebellum segmentation approach

The training process diagram for segmenting the cerebellum using encoder-decoder-based models is shown in Fig. 6, which contains mainly two phases: 1) Data Collection and Preprocessing, and 2) Training of the encoder-decoder-based model to segment the cerebellum of the fetal brain from 2D ultrasound images. Fig. 7 depicts the testing process for the proposed cerebellum segmentation approach.

4.3. Evaluation parameters

During the experiments, the effectiveness of cerebellum segmentation is evaluated using three metrics: the Dice Coefficient (DC), Intersection over Union (IoU), and Precision (PR), as outlined in the next equations

$$\begin{aligned}
 (1) \quad DC &= \frac{2 \cdot |P \cap G|}{|P| + |G|}, \\
 (2) \quad IoU &= \frac{|P \cap G|}{|P \cup G|}, \\
 (3) \quad PR &= \frac{|P \cap G|}{|P|},
 \end{aligned}$$

where: P is Predicted mask; G is Ground truth mask; $|P|$ is number of foreground pixels in the predicted mask; $|G|$ is number of foreground pixels in the ground truth mask; $|P \cap G|$ is number of overlapping foreground pixels; $|P \cup G|$ is number of pixels in either P or G .

4.4. Impact of batch sizes

By evaluating three configurations of batch size, namely 8, 16, and 32, we examine the effects of varying batch sizes on the performance of the proposed cerebellar segmentation techniques. A quantitative comparison of the proposed cerebellar segmentation techniques for various batch sizes is shown in Table 4. During this experiment, the optimizer is Adam, and the loss function is Binary Cross-Entropy.

The experimental results show that, for a relatively small dataset, a batch size of 8 outperforms all U-Net variants by enabling more frequent parameter updates, thereby improving the model's generalization.

Table 4. Quantitative comparison of segmentation results across different batch sizes

Approach	Batch size	Training				Validation			
		Loss	Dice coefficient	IoU	Precision	Loss	Dice coefficient	IoU	Precision
U-Net	8	7.28	72.32	56.45	79.51	6.89	70.03	53.67	73.23
	16	8.15	65.92	49.78	66.71	8.90	63.05	46.19	64.73
	32	11.01	61.89	44.56	64.81	11.56	58.00	40.91	59.86
Attention-U-Net	8	4.56	79.41	65.85	83.67	4.89	75.09	60.11	80.47
	16	5.71	73.07	57.71	75.91	6.78	72.11	56.28	75.89
	32	5.09	72.12	55.98	74.90	5.82	69.83	53.52	72.77
VGG16-U-Net	8	6.10	79.77	66.72	82.54	8.52	72.50	56.94	78.04
	16	5.83	77.81	63.15	80.08	4.90	69.99	53.32	72.85
	32	6.25	72.05	56.10	74.76	5.01	65.18	48.42	68.13
ResNet50-U-Net	8	3.04	84.90	73.49	85.26	9.12	83.74	72.52	84.05
	16	5.97	80.88	67.79	81.01	12.48	75.11	60.92	80.93
	32	5.13	81.37	68.41	80.45	15.05	79.86	66.49	83.70
EfficientNetB1-U-Net	8	2.56	89.52	81.06	89.83	2.07	92.38	85.46	93.03
	16	6.27	86.99	76.95	78.92	7.20	81.40	68.19	85.97
	32	5.81	78.35	64.82	80.76	6.55	68.76	52.58	84.01

4.5. Effect of optimizers

By evaluating two configurations of optimizers, namely Adam and RMSprop, we examine the effects of varying optimizers on the performance of the proposed cerebellar segmentation techniques. Table 5 presents an equivalence comparison of the two optimizers, Adam and RMSprop, for the proposed cerebellar segmentation techniques. This experiment uses the binary cross-entropy loss function with a batch size of 8.

Table 5. An equivalence comparison of segmentation results across different optimizers

Approach	Optimizer	Training				Validation			
		Loss	Dice coefficient	IoU	Precision	Loss	Dice coefficient	IoU	Precision
U-Net	Adam	7.28	72.32	56.45	79.51	6.89	70.03	53.67	73.23
	RMSprop	9.24	66.03	49.17	71.86	10.36	61.67	44.51	69.07
Attention-U-Net	Adam	4.56	79.41	65.85	83.67	4.89	75.09	60.11	80.47
	RMSprop	5.93	78.63	64.77	81.91	6.11	72.56	56.53	77.42
VGG16-U-Net	Adam	6.10	79.77	66.72	82.54	8.52	72.50	56.94	78.04
	RMSprop	7.75	80.97	68.21	82.78	7.81	71.81	56.43	75.06
ResNet50-U-Net	Adam	3.04	84.90	73.49	85.26	9.12	83.74	72.52	84.05
	RMSprop	4.83	80.83	67.39	82.00	6.98	75.61	60.11	78.56
U-Net-EfficientNetB1	Adam	2.56	89.52	81.06	89.83	2.07	92.38	85.46	93.03
	RMSprop	2.97	81.91	69.39	83.08	4.88	78.35	64.87	85.19

A conclusion from the experimental result is that the Adam optimizer surpassed the RMSprop optimizer in the context of loss.

4.6. Impact of loss functions

By evaluating three configurations of loss functions, namely Binary-Cross-Entropy loss (BCE loss), Dice loss, and Tversky loss, as defined in the following equations, respectively, we examine the effects of various loss functions on the performance of the proposed cerebellar segmentation techniques,

$$(4) \quad \text{BCE loss} = -\frac{1}{N} \sum_{i=1}^N [S_i \log(\hat{S}_i) + (1 - S_i) \log(1 - \hat{S}_i)],$$

$$(5) \quad \text{Dice loss} = 1 - \frac{2 \cdot \sum_{i=1}^N S_i \hat{S}_i + \varepsilon}{\sum_{i=1}^N S_i + \sum_{i=1}^N \hat{S}_i + \varepsilon}$$

$$(6) \quad \text{Tversky loss} = 1 - \frac{\sum_{i=1}^N S_i \hat{S}_i + \varepsilon}{\sum_{i=1}^N S_i \hat{S}_i + \alpha \sum_{i=1}^N S_i (1 - \hat{S}_i) + \beta \sum_{i=1}^N (1 - S_i) \hat{S}_i + \varepsilon},$$

where: N = Output size; $\hat{S}_i$ = i -th scalar value predicted by the model; S_i = corresponding ground truth label; α controls the penalty for false negatives; β controls the penalty for false positives; ε is a small constant added to avoid division by zero.

Moreover, we examine the performance of the proposed cerebellum segmentation approaches that combine the Dice and Tversky losses with the BCE loss function, namely BCE + Dice loss and BCE + Tversky loss.

The results of cerebellar segmentation using multiple models with different loss functions, such as U-Net, Attention-U-Net, VGG16-U-Net, ResNet50-U-Net, and EfficientNetB1-U-Net, are shown in Table 6.

Table 6. An equivalence comparison of segmentation results across different loss functions

Approach	Loss Function	Training				Validation			
		Loss	Dice coefficient	IoU	Precision	Loss	Dice coefficient	IoU	Precision
U-Net	BCE	7.28	72.32	56.45	79.51	6.89	70.03	53.67	73.23
	Dice	8.20	91.51	84.65	90.45	9.26	90.69	82.96	89.45
	Tversky	7.34	89.09	80.31	89.40	7.47	88.12	77.98	89.31
	BCE + Dice	3.00	91.78	84.78	93.95	3.11	90.89	83.51	95.43
	BCE + Tversky	1.45	93.20	87.67	93.34	1.40	91.23	84.68	92.02
Attention-U-Net	BCE	4.56	79.41	65.85	83.67	4.89	75.09	60.11	80.47
	Dice	5.67	94.20	89.37	95.45	6.06	93.49	87.09	92.03
	Tversky	4.50	71.89	56.38	72.05	1.77	67.76	51.39	68.93
	BCE + Dice	1.03	95.61	91.39	95.95	1.97	94.25	89.35	95.02
	BCE + Tversky	0.23	93.34	87.74	96.10	2.47	91.67	84.22	95.29
VGG16-U-Net	BCE	6.10	79.77	66.72	82.54	8.52	72.50	56.94	78.04
	Dice	2.88	96.17	94.63	90.34	4.65	94.91	90.99	95.52
	Tversky	3.73	84.72	73.19	86.06	4.05	82.63	70.46	85.43
	BCE + Dice	7.40	97.66	93.43	98.51	7.21	94.99	90.57	96.82
	BCE + Tversky	2.01	94.95	90.39	96.03	2.94	92.85	86.83	97.41
ResNet50-U-Net	BCE	3.04	84.90	73.49	85.26	9.12	83.74	72.52	84.05
	Dice	5.31	93.47	87.48	94.31	5.84	92.42	85.83	93.40
	Tversky	7.04	84.03	72.47	86.92	7.51	81.97	69.05	84.03
	BCE + Dice	0.53	94.31	89.25	97.26	0.55	93.38	87.09	95.20
	BCE + Tversky	0.17	95.48	91.36	98.54	0.24	95.72	91.31	95.98
EfficientNetB1-U-Net	BCE	2.56	89.52	81.06	89.83	2.07	92.38	85.46	93.03
	Dice	8.93	91.08	83.63	92.05	9.03	89.73	81.95	90.44
	Tversky	4.77	93.26	87.38	94.61	4.93	90.70	82.96	90.93
	BCE + Dice	1.52	96.81	94.37	98.89	1.72	92.65	86.31	96.75
	BCE + Tversky	0.09	97.82	95.79	98.01	0.11	95.94	93.29	98.42

The first experiment evaluates the U-Net's performance in segmenting the cerebellum. The experimental results are represented graphically in Fig. 8. The

binary cross-entropy loss function combined with the Tversky loss function yields the lowest training loss of 1.45% and validation loss of 1.40%, as shown in Fig 8a and b. Additionally, Fig 8c and d show that the maximum training Dice coefficient of 93.20% and validation Dice coefficient of 91.23% are obtained when the binary cross-entropy loss function and the Tversky loss function are combined. In terms of IoU, Fig 8e and f demonstrate how the binary cross-entropy loss function plus the Tversky loss function works better than other loss functions. In particular, IoUs of 87.67% for training and 84.68% for validation are reported. Nevertheless, experimental findings indicate that higher Dice and IoU scores are achieved with the combination of the binary cross-entropy loss function and the Tversky loss function, but not with the combination of the binary cross-entropy loss function and the Dice loss function. Specifically, the BCE + Dice loss function yields a training precision of 93.45% and a validation precision of 95.43%.

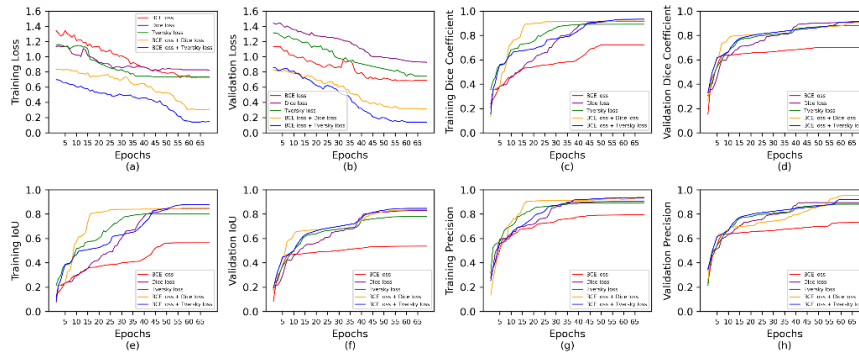


Fig. 8. A comparison of the training and validation experimental outcomes with five different loss functions for cerebellum segmentation using U-Net

The Attention-U-Net is assessed for cerebellar segmentation in the second experiment. Fig. 9 shows the findings of the cerebellum segmentation experiment graphically. Fig. 9a illustrates the lowest training loss of 0.23% when the binary cross-entropy loss function and the Tversky loss function are combined. However, as shown in Fig. 9b, the Tversky loss function achieves the lowest validation loss of 1.77%. Conversely, the combination of the dice loss function with the binary cross-entropy loss function yielded the greatest training dice coefficient of 95.61% and validation dice coefficient of 94.25%. Additionally, the greatest training IoU of 91.39% and validation IoU of 89.35% are obtained when the dice loss and binary cross-entropy loss function are combined. The cerebellum is segmented more accurately when the binary cross-entropy loss function and the Tversky loss function are combined, as shown in Fig 9g and h, with training precision of 96.10% and validation precision of 95.29%.

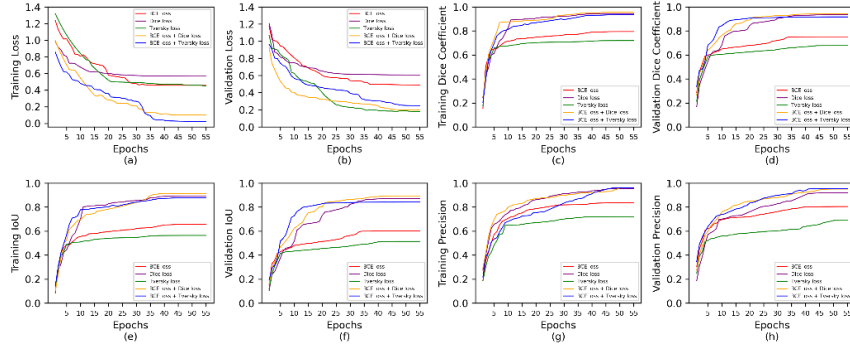


Fig. 9. A comparison of the training and validation experimental outcomes with five different loss functions for cerebellum segmentation using Attention-U-Net

In the third experiment, we evaluated three U-Net methods based on transfer learning: VGG16-U-Net, ResNet50-U-Net, and EfficientNetB1-U-Net. The U-Net's contraction path in this experiment includes the use of pre-trained weights. The VGG16-U-Net graphical findings are shown in Fig. 10. The binary cross-entropy loss function combined with the Tversky loss function yields the lowest training loss of 2.01% and validation loss of 2.94%, as shown in Fig 10a and b. Furthermore, the highest training Dice coefficient of 97.66% and validation Dice coefficient of 94.99% are obtained by combining the binary cross-entropy loss function with the Dice loss function; Fig 10e and f show that the Dice loss function achieves the highest training and validation IoU values. However, training the model with a combination of binary cross-entropy and Dice loss functions yields the highest training precision of 98.51%, while training with a combination of binary cross-entropy and Tversky loss functions yields the highest validation precision of 97.41%.

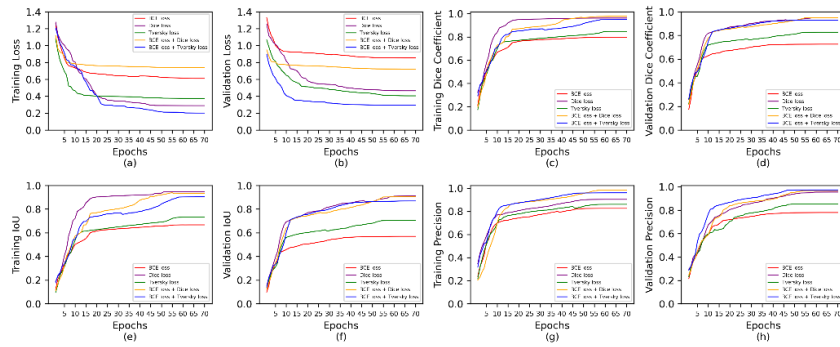


Fig. 10. A comparison of the training and validation experimental outcomes with five different loss functions for cerebellum segmentation using VGG16-U-Net

Fig. 11 shows a visual representation of the experimental results utilizing ResNet50-U-Net. The combination of the binary cross-entropy loss function and the

Tversky loss function yields the lowest training loss of 0.17%, the lowest validation loss of 0.24%, the highest training dice coefficient of 95.48%, the highest validation dice coefficient of 95.72%, the highest training IoU of 91.36%, the highest validation dice coefficient of 91.31%, the highest training precision of 98.54%, and the highest validation precision of 95.98%, respectively.

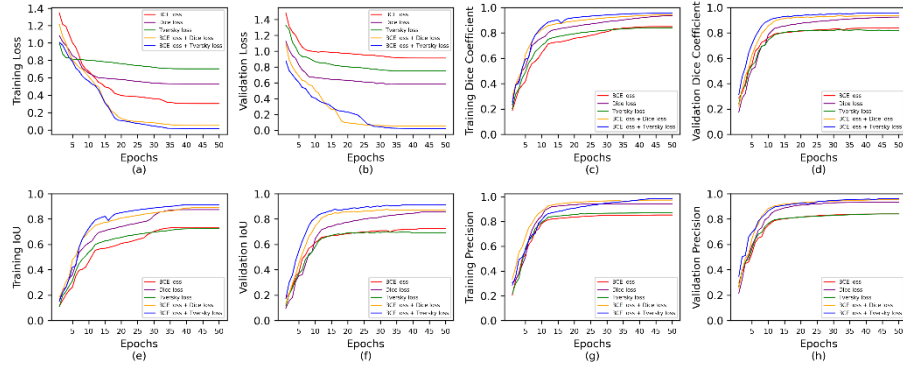


Fig. 11. A comparison of the training and validation experimental outcomes with five different loss functions for cerebellum segmentation using ResNet50-U-Net

The experimental findings of the cerebellum segmentation using EfficientNetB1-U-Net are shown graphically in Fig. 12. As seen in Fig 12a and b, the binary cross-entropy loss function combined with the Tversky loss function yields the lowest training loss of 0.09% and validation loss of 0.11%. Combining the binary cross-entropy loss function with the Tversky loss function results in the highest training dice coefficient of 97.82%, the highest validation dice coefficient of 95.94%, the highest training IoU of 95.79%, and the highest validation IoU of 93.29%.

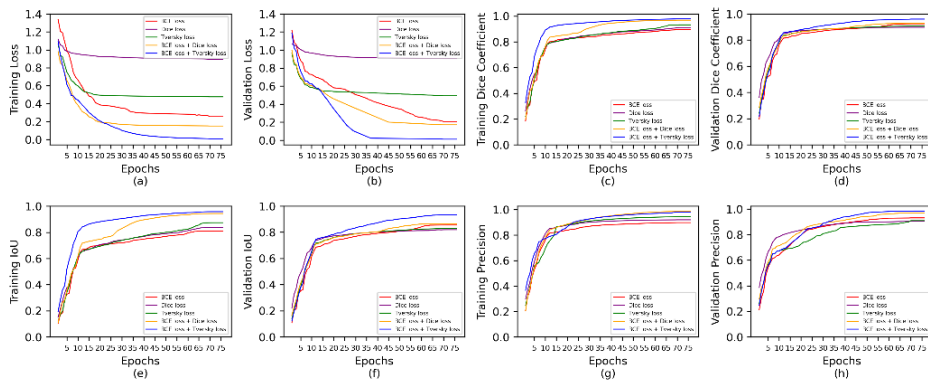


Fig. 12. A comparison of the training and validation experimental outcomes with five different loss functions for cerebellum segmentation using EfficientNetB1-U-Net

Additionally, Fig. 12g demonstrates that combining the binary cross-entropy and dice loss functions yields the maximum training precision of 98.89%. Conversely, Fig. 12h shows that combining the binary cross-entropy loss with the Tversky loss yields the highest validation precision of 98.42%.

Fig. 13 illustrates the binary segmentation images generated from the trained models. The comparative results demonstrate that Attention-U-Net generates less noise and segments more precise cerebellum boundaries. Additionally, U-Net segments precise cerebellum boundaries, though with more noise than Attention-U-Net. Nevertheless, some noise is produced in the segmentation binary image by VGG16-U-Net, ResNet50-U-Net, and EfficientNetB1-U-Net. Despite this, U-Net techniques based on transfer learning performed better during training and validation. Compared to VGG16-U-Net, ResNet50-U-Net, and EfficientNetB1-U-Net, U-Net and Attention-U-Net produce more accurate and precise binary segmentation results while producing less noise.

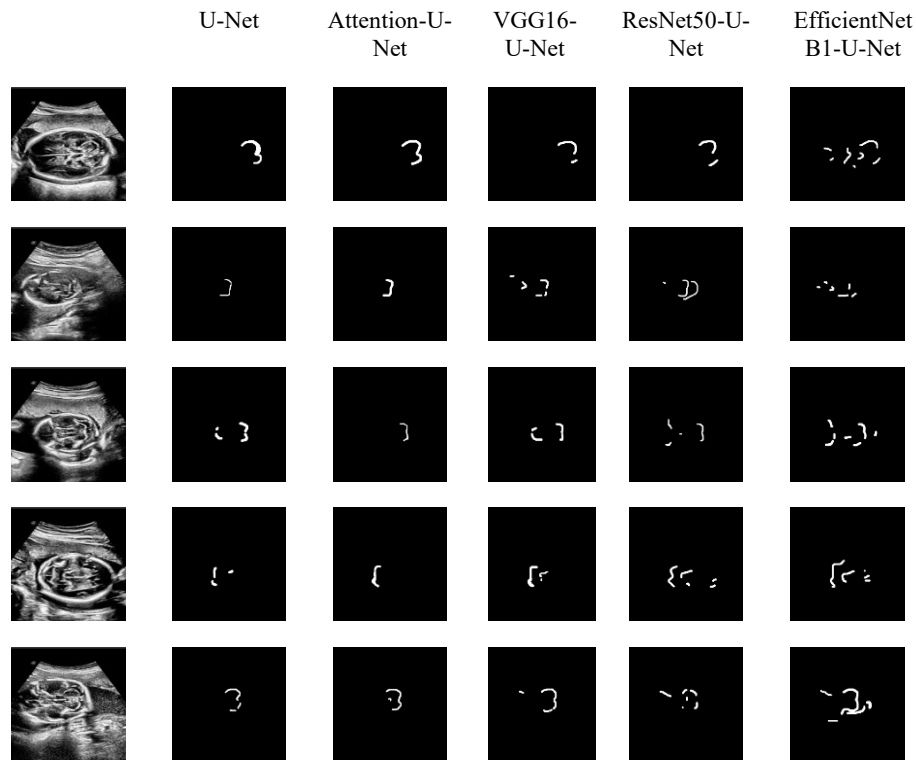


Fig. 13. Comparison analysis of the results obtained from the proposed cerebellum segmentation approaches

5. Conclusion

Segmentation helps identify the region of Interest, minimizing unnecessary information and leading to faster inference. This paper presents cerebellum segmentation using U-Net, Attention-U-Net, and three transfer-learning-based U-Net encoders: VGG16-U-Net, ResNet50-U-Net, and EfficientNetB1-U-Net. These cerebellum segmentation approaches are evaluated using three different batch sizes: 8, 16, and 32; two different optimizers: Adam and RMSprop; and three different loss functions: binary cross-entropy loss (BCE), dice loss, and Tversky loss, as well as combined loss functions BCE + dice loss and BCE + Tversky loss.

Though the experimental findings demonstrate the superior performance of transfer-learning-based U-Net techniques in training and validation, U-Net-based techniques performed better in testing results. Particularly, Attention-U-Net performed best among U-Net variants in our experiments. Additionally, combining BCE with other loss functions improves segmentation performance compared to binary cross-entropy, Dice, and Tversky losses.

In the future, we aim to extend our segmentation framework with a larger dataset and multi-view images. Moreover, we also aim to use more advanced segmentation techniques. In particular, we aim to use generative adversarial networks for automatic cerebellar segmentation and compare the outcomes with cerebellar segmentation methods utilizing U-Net and its variants.

References

1. Manzo, L., G. Rizzo, M. E. Pietrolucci, A. Capelli, L. Caforio, F. D'Antonio. Fetal Cerebellar Area: Ultrasound Reference Ranges at 13-39 Weeks of Gestation. – J. Clin. Med., Vol. **12**, 2023, No 12, 4080.
2. Manto, M., J. M. Bower, A. B. Conforto, J. M. Delgado-García, S. N. F. da Guarda, M. Gerwig et al. Consensus Paper: Roles of the Cerebellum in Motor Control – the Diversity of Ideas on Cerebellar Involvement in Movement. – Physio Rev., Vol. **92**, 2012, No 3, pp. 961-1025.
3. Anwar, S., S. Samin, A. K. Jadoon, R. A. Khan. Delving into Transfer Learning within U-Net for Refined Retinal Vessel Segmentation: An Extensive Hyperparameter Analysis. – Biomedical Signal Processing and Control, Vol. **87**, 2024, 105381.
4. Malinger, G., S. Ginath, T. Lerman-Sagie, N. Watemberg, D. Lev, M. Glezerman. The Fetal Cerebellar Vermis: Normal Development as Shown by Transvaginal Ultrasound. – Prenat Diagn, Vol. **21**, 2001, No 8, pp. 687-692.
5. Minaee, S., Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, D. Terzopoulos. Image Segmentation Using Deep Learning: A Survey. – IEEE Trans. Pattern Anal. Mach. Intell., Vol. **44**, 2021, No 7, pp. 3523-3542.

6. Garel, C., C. Fallet-Bianco, L. Guibaud. The Fetal Cerebellum: Development and Common Malformations. – *J. Child Neurol*, Vol. **26**, 2011, No 12, pp. 1483-1492.
7. Stoodley, C. J., C. Limperopoulos. Structure-Function Relationships in the Developing Cerebellum: Evidence from Early-Life Cerebellar Injury and Neurodevelopmental Disorders. – *Semin Fetal Neonatal Med*, Vol. **21**, 2016, No 5, pp. 356-364.
8. Xu, F., X. Ge, Y. Shi, Z. Zhang, Y. Tang, X. Lin et al. Morphometric Development of the Human Fetal Cerebellum during the Early Second Trimester. – *Neuroimage*, Vol. **207**, 2020, 116372.
9. Van der Linden, A., J. Sijbers, P. Scheunders, M. Verhoye, D. Van Dyck, E. Raman. Watershed-Based Segmentation of 3D MR Data for Volume Quantization. – *Magn Reson Imaging*, Vol. **15**, 1997, No 6, pp. 679-688.
10. Powell, S., V. A. Magnotta, H. Johnson, V. K. Jammalamadaka, R. Pierson, N. C. Andreasen. Registration and Machine Learning-Based Automated Segmentation of Subcortical and Cerebellar Brain Structures. – *Neuroimage*, Vol. **39**, 2008, No 1, pp. 238-247.
11. Vishal, S., S. Singh, N. Sharma, P. K. Gupta. Semantic Segmentation of Cerebellum in 2D Fetal Ultrasound Brain Images Using Convolutional Neural Networks. – *IEEE Access*, Vol. **9**, 2021, pp. 85864-85873.
12. Gangopadhyay, T., S. Halder, P. Dasgupta, K. Chatterjee, D. Ganguly, S. Sarkar, S. Roy. MTSE U-Net: An Architecture for Segmentation and Prediction of Fetal Brain and Gestational Age from MRI of Brain. – *Netw. Model Anal. Health Inform. Bioinform.*, Vol. **11**, 2022, No 1, 50.
13. Li, Y., B. Khanal, B. Hou, A. Alansary, J. J. Cerrolaza, M. Sinclair, D. Rueckert. Standard Plane Detection in 3D Fetal Ultrasound Using an Iterative Transformation Network. – In: *Proc. of International Conference on Medical Image Computing and Computer-Assisted Intervention*, Cham, Springer, 2018, pp. 392-400.
14. Chen, H., L. Wu, Q. Dou, J. Qin, S. Li, J.-Z. Cheng, D. Ni, P.-A. Heng. Ultrasound Standard Plane Detection Using a Composite Neural Network Framework. – *IEEE Trans. Cybern.*, Vol. **47**, 2017, No 6, pp. 1576-1586.
15. Jang, J., Y. Park, B. Kim, S. M. Lee, J. Y. Kwon, J. K. Seo. Automatic Estimation of Fetal Abdominal Circumference from Ultrasound Images. – *IEEE J. Biomed. Health Inform.*, Vol. **22**, 2018, No 5, pp. 1512-1520.
16. Thomas, J. G., R. A. Peters, P. Jeanty. Automatic Segmentation of Ultrasound Images Using Morphological Operators. – *IEEE Trans. Med. Imaging*, Vol. **10**, 1991, No 2, pp. 180-186.
17. Van den Heuvel, T. L. A., D. de Bruijn, C. de Korte, B. van Ginneken. Automated Measurement of Fetal Head Circumference Using 2D Ultrasound Images. – *PLoS One*, Vol. **13**, 2018, No 8, e0200412.
18. Coronado González, E., R. Valdés Rodríguez, C. R. Robbinc, M. Pinto, D. Salas González. Convolutional Neural Networks for Segmenting Cerebellar Fissures from Magnetic Resonance Imaging. – *Sensors*, Vol. **22**, 2022, No 4, 1345.
19. Ronneberger, O., P. Fischer, T. Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. – In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Springer, Cham, 2015, pp. 234-241.

20. Tustison, N. J., B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich, J. C. Gee. N4ITK: Improved N3 Bias Correction. – IEEE Trans Med Imaging, Vol. **29**, 2010, No 6, pp. 1310-1320.
21. Fonov, V. S., A. C. Evans, R. C. McKinstry, C. R. Almlie, D. L. Collins. Unbiased Nonlinear Average Age-Appropriate Brain Templates from Birth to Adulthood. – Neuroimage, Vol. **47**, 2009, Suppl. 1, S102.
22. Bolduc, M. E., A. J. du Plessis, N. Sullivan, N. Guizard, X. Zhang, R. L. Robertson, C. Limperopoulos. Regional Cerebellar Volumes Predict Functional Outcome in Children with Cerebellar Malformations. – Cerebellum, Vol. **11**, 2012, No 2, pp. 531-542.
23. Lee, D. K., U. Yoon, K. Kwak, J. M. Lee. Automated Segmentation of Cerebellum Using Brain Mask and Partial Volume Estimation Map. – Comput. Math. Methods Med., Vol. **2015**, 167489.
24. Walsh, J., A. Othmani, M. Jain, S. Dev. Using U-Net Network for Efficient Brain Tumor Segmentation in MRI Images. – Healthcare Analytics, Vol. **2**, 2022, 100098.
25. Ali, M., S. O. Gilani, A. Waris, K. Zafar, M. Jamil. Brain Tumour Image Segmentation Using Deep Networks. – IEEE Access, Vol. **8**, 2020, pp. 153589-153598.
26. Chauhan, H. N., C. B. Thacker. A Novel Hybrid Deep Learning Framework for Image Captioning Using Combined Semantic and Object-Level Features. – Cybernetics and Information Technologies, Vol. **25**, 2025, No 4, pp. 116-128.
27. Pravitasari, A. A., N. Iriawan, M. Almuhayar, T. Azmi, Irhamah, K. Fithriasari, S. W. Purnami, W. Ferriastuti. UNet-VGG16 with Transfer Learning for MRI-Based Brain Tumor Segmentation. – In: Journal of Physics: Conference Series. Vol. **1477**. IOP Publishing, 2020, 022014.
28. Yu, W., J. Shen, G. Ning, Y. Kan, J. Lin, J. He, L. Luo. Automatic Classification of Leukocytes Using Deep Neural Network. – In: Proc. of 12th IEEE International Conference on ASIC (ASICON'17), IEEE, 2017, pp. 1041-1044.
29. Siddique, N., S. Paheding, C. P. Elkin, V. Devabhaktuni. U-Net and Its Versions for Medical Image Segmentation: A Review of Theory and Applications. – IEEE Access, Vol. **9**, 2021, pp. 82031-82057.
30. Oktay, O., J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa et al. Attention U-Net: Learning where to Look for the Pancreas. – arXiv preprint, 2018.
31. Schlemper, J., O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, D. Rückert. Attention-Gated Networks: Learning to Leverage Salient Regions in Medical Images. – Medical Image Analysis, Vol. **53**, 2019, pp. 197-207.
32. Mecifi, Y. Z., M. Merzoug, A. Etchiali, M. M'hamedi, F. Hadjila, A. Bekkouche. Early Skin Cancer Detection Using the SLICE-3D Dataset: A Transfer Learning Model and DCGAN Approach to Address Data Imbalance. – Cybernetics and Information Technologies, Vol. **25**, 2025, No 3, pp. 39-53.
33. Ramirez Zegarra, R., T. Ghi. Use of Artificial Intelligence and Deep Learning in Fetal Ultrasound Imaging. – Ultrasound Obstet Gynecol, Vol. **62**, 2023, No 2, pp. 185-194.
34. Hesse, L. S. Quantification of Fetal Brain Development from Ultrasound Images Using Interpretable Deep Learning. PhD Thesis, University of Oxford, 2023.
35. D'Angelo, E. Physiology of the Cerebellum. – In: Handbook of Clinical Neurology. Vol. **154**. 2018, pp. 85-108.

36. Lu, G., J. Luo, X. Zhou, L. Shao, S. Xu, M. Xu. ECAU-Net: Efficient Channel Attention U-Net for Fetal Ultrasound Cerebellum Segmentation. – Computerized Medical Imaging and Graphics, Vol. **92**, 2021, 101967.

Received: 25.03.2026; Accepted: 01.06.2026

© 2026 by the authors. Licensee IICT-BAS, Sofia, Bulgaria. This article is an open access article distributed under the terms and conditions of the **Creative Commons Attribution-NonCommercial-NoDerivatives** license.