Proceedings of the Second International Conference on Automation, Computing and Renewable Systems (ICACRS-2023)
IEEE Xplore Part Number: CFP23CB5-ART; ISBN: 979-8-3503-4023-5

2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS) | 979-8-3503-4023-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICACRS58579.2023.10404359

Enhancing Shift-Invariance for Accurate Brain MRI Skull-Stripping using Adaptive Polyphase Pooling in Modified U-Net

Adish Padalia
Department of Information Technology
Sardar Patel Institute of Technology
Mumbai, India
padaliaadish@gmail.com

Tanish Shah
Department of Information Technology
Sardar Patel Institute of Technology
Mumbai,India
tanish.shah07@icloud.com

Pratik Pujari
Department of Computer Engineering
Sardar Patel Institute of Technology
Mumbai, India
pratikpujari1000@gmail.com

Dr. Aarti Karande
Department of Computer Science and Engineering
Sardar Patel Insitute of Technology
Mumbai, India
aartimkarande@spit.ac.in

***Abstract*—Image segmentation is an essential aspect of image processing, where regions of interest (ROI) are identified by employing feature descriptors like edges, color, and texture. Medical images often contain unwanted segments, necessitating preprocessing steps to extract valuable information. In the realm of medical image segmentation, skull-stripping in 3D volumetric Brain MRI images holds significant importance. This study delves into supervised learning techniques to extract crucial brain features from non-brain tissues. The approach involves a modified U-net that operates on 2D slices of the brain MRI, yielding a skull-stripped image with superior accuracy compared to existing methods. The architecture incorporates Adaptive Polyphase Pooling layers, enhancing the architecture's robustness and the model's generalization. These Adaptive Polyphase Pooling layers ensure shift-invariance in the predicted output, maintaining consistency even if the single view of the Brain MRI shifts along the two axes in the plane.**

***Index Terms*—Adaptive Polyphase Pooling, U-net, Supervised-learning, skull-stripping, Brain Magnetic Resonance Imaging**

I. Introduction

Image processing is a method used to transform images in order to obtain useful and relevant information from it. It is widely used in many instances of our daily life and in the professional field in the form of image enhancement, compress,ion and removal of noise [1]. Medical images are contaminated with unwanted segments and thus need some pre-processing steps to extract beneficial information. By applying pre-processing steps on a medical image, it can help to analyse it better and can improve the chances of detecting illness effectively [2]. Image segmentation extracts the areas referred to as the Region of Interest (ROI) through feature descriptors like edges, color and texture in a given image. To enhance the performance of segmentation and feature extraction techniques, image pre-processing attempts to increase the quality of the images present in a dataset [3].One of the uses of segmentation is in analyzing brain development, image in Magnetic Resonance Imaging (MRI). Image segmentation will help in divide the areas of brain tissue into white matter (WM), gray matter (GM) and cerebrospinal fluid (CSF). Segmentation machine learning method is mainly classified into two groups, supervised learning and unsupervised learning. MRI scans produce 3D brain images, necessitating the use of computer-aided tools for tasks such as diagnosis and detailed anatomical analysis. Image processing [4], particularly segmentation, is critical in precisely isolating and comprehending various brain structures for medical image analysis. The emphasis in image segmentation has traditionally been on 2D images. When working with 3D volumetric data, such as from MRI scans, the process entails slicing the 3D data into individual 2D images, with segmentation applied sequentially to each slice. Although 2D segmentation methods can be used in 3D space, the overall time complexity and computation time increases.

II. Related Work

A. *Supervised Learning*

An overview of supervised learning's use in medical image processing is given in this work [5], with particular attention on segmentation, detection, and classification. It provides an overview of the existing medical imaging datasets for a range of disorders and describes the performance measures of supervised learning. For each job, it also examines a few of the most well-liked and significant deep learning techniques, including ResNet, VGGNet, and GoogLeNet for classification, Faster R-CNN, YOLO, and Mask R-CNN for detection, and FCN, U-Net, and DeepLab for segmentation. The difficulties and potential paths of supervised learning for medical image processing are covered.

The most recent self-supervised learning techniques from the field of computer vision are covered in these publications as they relate to medical imaging analysis [6]. Self-supervised learning is a subset of unsupervised learning in which learning tasks or labels are generated by the data itself. The drawbacks of supervised learning, such as the requirement for substantial quantities of labelled data and the inability to generalise to new domains, can be addressed by self-supervised learning. The papers discuss the applications of self-supervised learning methods in medical image analysis tasks, including segmentation, classification, registration, reconstruction, and synthesis, and they classify the methods as predictive, generative, and contrastive approaches. The articles also cover the difficulties and potential paths for self-supervised learning in the field of medical imaging analysis.

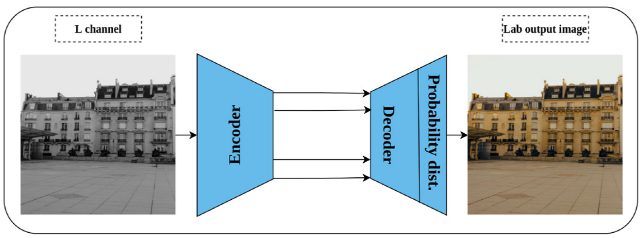

Fig. 1. An example of image denoising for the learning of self-supervised features [6].

In this study [7], a computer-aided design (CAD) system is proposed to aid in the early identification and detection of glaucoma, an irreversible blindness-causing condition. Three pre-trained convolutional neural networks (CNNs) are combined in the system's ensemble-based deep learning model: ResNet, VGGNet, and GoogLeNet. Fundus pictures are sent into the system, which generates a binary label indicating the whether glaucoma is present or not. The DRISHTI-GS, DRIONS-DB, HRF, PSGIMSR, and a combined dataset are the five datasets on which the system demonstrates good accuracy, sensitivity, and specificity.

In order to address a variety of medical image analysis problems, including segmentation, classification, detection, registration, reconstruction, synthesis, and enhancement, this study examines and summarises current works that use deep learning approaches [8]. A few clinical uses of deep learning in medical image processing, including diagnosis, prognosis, treatment planning, and monitoring, are also covered in the study. The study outlines the benefits and difficulties of deep learning in the context of medical picture analysis and offers some ideas and recommendations for further investigation.

Table.I shows an elaborative comparison of the advantages and limitations of unsupervised learning-related work.

TABLE I
COMPARISON OF SUPERVISED LEARNING RELATED WORK

Authors	Merits	Demerits
A. Aljuaid and M. Anwar [5]	Comprehensive analysis of popular deep learning techniques.	Limited discussion on specific techniques and focuses on current datasets, ignores potential newer ones
S. Joshi et al [7]	Early identification of glaucoma using a computer-aided design (CAD) system.	Reliance on an ensemble-based deep learning model offers little insight into the possible drawbacks of the proposed system
S. Shurrab and R. Duwairi [6]	Coverage of recent self-supervised learning strategies in medical imaging.	Does not provide a thorough breakdown of each strategy and lacks detailed discussion on real-world implementation issues
X. Chen et al [8]	Summarizes recent research on deep learning in medical image processing.	Does not provide information for examining certain clinical uses in great depth

B. Unsupervised Learning

In the research conducted by Wilfried-Woeber et.al [9], an examination of supervised and unsupervised learning techniques for natural image processing was undertaken. The paper begins by tracing the evolution of feature extraction, from manual to automated methods through CNNs. It introduces the application of geometric morphometrics in biodiversity research and the role of machine learning in reducing uncertainty in landmark positions. The study further evaluates various unsupervised approaches, including PCA, ICA, Autoencoders, and GPLVM, for their effectiveness in dimensionality reduction and latent representation extraction in image analysis.

Yamawaki K et.al [10] introduces a blind image super-resolution (SR) technique utilizing deep unsupervised learning, aiming to enhance the quality of low-resolution images without external training data or prior knowledge. The method employs a generative encoder-decoder network to learn high-resolution image features and a learnable depth-shared convolutional layer to estimate degradation factors. It is trained end-to-end, minimizing a loss function involving reconstruction and perceptual errors. The results show superior performance compared to bicubic interpolation and competitive results with other SR methods, even under diverse imaging conditions with unknown degradation models. This approach demonstrates promise for real-world super-resolution applications, offering a robust and effective solution.

López et al [11] present a comprehensive evaluation of unsupervised MRI denoising methods, introducing innovative approaches—Stein's Unbiased Risk Estimator (SURE) and a blindspot network. These methods are tested on real knee MRI and synthetic brain MRI datasets, outperforming the state-of-the-art Non-Local Means (NLM) algorithm in both quantitative and qualitative assessments. By addressing the need for unsupervised approaches in MRI denoising, their research opens new possibilities for versatile and robust denoising strategies. The study underscores the potential of these methods, with the blindspot network exhibiting a slight edge over the SURE model, while acknowledging considerations related to over-training and noise assumptions.

Table.II makes a comparative analysis of different unsupervised learning-related works and shows the merits and demerits of each research paper.

TABLE II
COMPARISON OF UNSUPERVISED LEARNING RELATED WORK

Authors	Merit	Demerits
Wilfried-Woeber et.al [9]	Comprehensive review of popular brain MRI segmentation methods, covering techniques, differences, capabilities, and limitations.	Limited focus on MRI segmentation of the human brain, potentially excluding coverage of other medical image analysis aspects.
Yamawaki K et.al [10]	The paper presents a comprehensive review of existing skull stripping methods, categorizing them into five different categories and highlights the importance of skull stripping.	The majority of existing skull stripping methods are designed for T1-weighted brain images and may not be applicable to all brain image types and orientations.
López et al [11]	Evaluates and compares different MRI denoising methods using unsupervised learning and provides a comprehensive analysis of the performance of these methods.	Evaluation is based on simulated data rather than real patient data and does not provide a direct comparison with other existing denoising methods.

III. BACKGROUND INFORMATION

A. Magnetic Resonance Imaging (MRI)

MRI is a non-invasive medical imaging technique that uses strong magnetic fields and radio waves to produce detailed 3-dimensional anatomical images of the human body, particularly the brain. MRI has the ability to image the brain in any plane without physically moving the patient whereas CT scans are limited to one plane, the axial plane [12]. Brain MRI images are essential in diagnosing neurological diseases such as Alzheimer's, childhood brain tumors, stroke, multiple sclerosis, and brain aneurysms [13].

B. Skull Stripping

Skull stripping, a crucial step in quantitative morphometric analysis of MR brain images, involves the differentiation of brain tissue from non-brain regions in MRI head scans [14]. This process is essential to ensure accurate results before implementing various image processing algorithms such as image registration and warping [15], brain volumetric measurement [16], inhomogeneity correction [17], analysis of cortical structure [18], identification of brain regions [19], multiple sclerosis studies [20], Alzheimer's disease research [21], and schizophrenia analysis [22]. This initial step is vital in preparing the data and achieving precise outcomes in subsequent analyses and investigations of neurological conditions and brain structures.

K. Somasundaram and P. Kalavathi [23] propose a fully automatic method for skull stripping in MRI head scans using 2D region growing in two stages In the first stage, the brain portion is extracted from the middle slice of the brain volume. This is done by using Otsu's automatic thresholding to obtain a binary image, which is then used to find the rough brain mask. The rough brain mask is further refined using the 2D region growing method to obtain the fine brain mask. A circle is defined inside the rough brain to select the seed points for region growing. In the second stage, the same procedure is applied to extract the brain portions in the remaining slices. The geometric similarities between adjacent slices are utilized to extract the brain portions in a hierarchical manner. The advantages of the proposed method include its ability to accurately extract the brain portion from T1, T2, and PD-weighted MR images. The method utilizes a combination of Otsu's automatic thresholding and 2D region growing to refine the rough brain mask and obtain a fine brain area. The geometric similarities of adjacent slices are also utilized to extract brain portions in the remaining slices. Experimental results demonstrate that the proposed method outperforms existing methods such as BET and BSE in terms of accuracy.

The research methodology used by C. Balaji and S. Veni [24] on automatic skull stripping from MRI images of the human brain reveals significant benefits and challenges. Using a deep learning framework, specifically the U-Net model, allows for precise brain tissue detection, demonstrating deep learning's efficacy in image analysis and computer vision. With a Dice score of 98%, the proposed modified U-Net model demonstrates commendable competency in extracting the brain portion from non-brain tissues. A useful aspect of the research is a comparison with other deformable and deep learning models, such as BET2, RoBEX, and UNet3D, which contributes to the evaluation of the proposed model's performance. However, challenges remain, such as the limited availability of relatively small skull-stripping datasets, which may have an impact on the deep learning model's training and generalization capabilities. Furthermore, the requirement for domain-specific pre-processing and the time-consuming nature of these stages, combined with the potential impact of data drift on model performance, necessitates additional thought and refinement for enhanced applicability in real-world scenarios.

C. Advantages of using Supervised Learning methods

Supervised learning is a powerful tool in machine learning that utilizes labeled data to train models capable of making accurate predictions and decisions. In the field of brain MRI image processing, supervised algorithms can accurately identify the location, shape, size, and type of brain tumors, enabling precise diagnosis and prognosis by analyzing MRI images.

- Through SL, MRI images can be divided into regions corresponding to various brain tissues like gray matter, white matter, and cerebrospinal fluid. This segmentation aids in measuring the volume and thickness of brain tissues, as well as detecting any abnormalities or changes over time
- Supervised learning is used in brain MRI image processing tasks to identify and classify different structures in the brain, such as tumors, blood vessels, and gray matter.
- It is preferred over unsupervised learning because it can provide more accurate results by using labeled data to

train the model, which can then be used to classify new, unseen data with high accuracy [25]

- Unsupervised learning, on the other hand, helps to find patterns and relationships in unlabeled data without any prior knowledge of the output [26].

D. Different Structures in Brain

Understanding the varied shades of grey that correspond to distinct tissues and structures is necessary to identify various brain regions using MRI pictures.

- Darker hues can be used to depict fluid-filled regions, such as the cerebrospinal fluid.
- Grey matter and other softer tissues are commonly shown in medium greys.
- Denser tissues, such as big blood arteries or white matter, are shown in lighter tones.

An organized method should be used to analyze a brain MRI, beginning at the midline and moving laterally. The ventricles are typically analysed first, followed by the surrounding subcortical tissues, brain lobes, cerebral cortex, meninges, and skull. A valuable resource for understanding anatomical structures including the diencephalon, mesencephalon (midbrain), pons, myelencephalon (medulla oblongata, bulb), and spinal cord is an MRI picture.

1) Brain Tissue Segmentation: Using labeled datasets, supervised learning algorithms divide MRI images into areas that represent various brain tissues, including white matter, grey matter, and cerebrospinal fluid (CSF). The algorithms are able to precisely identify these locations by utilizing unique properties that were recorded in these labeled pictures. This allows for the assessment of tissue volume and thickness as well as the detection of anomalies or changes over time.

2) Identification and Characterization: By using supervised learning, these algorithms are able to precisely recognize and categorize a variety of brain structures, such as tumours, blood vessels, and particular regions of grey matter. By training on labeled datasets that contain examples of these structures, these algorithms are able to pick up distinguishing characteristics that help with their characterization and identification.

3) Feature Extraction and Pattern Recognition: To identify minute distinctions between different brain areas, supervised learning systems use feature extraction and pattern recognition procedures. These techniques let the algorithms identify complex patterns and characteristics, which makes it easier to categorize various brain areas or abnormalities.

4) Integration of Multi-modal Data: To improve the precision of structure categorization, supervised learning models can incorporate data from a variety of imaging modalities, including as MRI, fMRI, and DTI. Through the integration of data from several imaging modalities, these models can offer a more thorough comprehension of intricate brain architectures and their relationships.

E. 2D Brain MRI Scans

Utilizing 2D slices of brain MRI instead of volumetric images in the model makes analyzing medical images using 2D slices a more straightforward method, preferred by medical professionals for its ease of interpretation and understanding. Annotating and reporting findings on 2D slices is commonly practiced and more straightforward, making it a popular method of analysis in medical reports. In clinical settings, radiologists and medical practitioners prefer to analyze medical images slice by slice, particularly for detailed assessments or when making specific diagnoses. This approach aligns with the established methodology and practices in medical imaging interpretation and is widely used for its simplicity and efficiency.

F. Lack of Shift Invariance in CNN

CNNs, commonly employed for image classification, incorporate fully connected layers toward the end of their architecture. Its job is to analyze medical pictures, especially MRI scans, for purposes including detecting grey and white matter in the brain and stripping skulls. Features are extracted from images and classified using CNN. Nevertheless, it lacks shift-invariance which indicates that even small shifts in the convolutional feature maps of the final layer could affect the output of the classifier. Ideally, an image classifier's output should not be affected by small shifts in the input image. It would be extremely difficult to maintain consistency in the predicted output for skull-stripping if shifts in the MRI image lead to changes in the output. These changes could have disastrous consequences on subsequent attempts to predict abnormalities or process the brain further in order to identify the key regions of the brain such as gray matter, white matter and so on. It was previously assumed that CNN's exhibit this desirable property [27]–[29] but current findings suggest that they are not consistent with the shifts. Several works have demonstrated this, including [30]–[32].

G. Partial Solution in adding Shift Invariance

Initially, researchers attempted to use max-pooling [33] and average pooling [28] methods to achieve shift-equivariance in the CNN architecture, but these techniques were unsuccessful. To make the CNN model shift-equivariant, an anti-aliasing method was introduced [32]. This method involves introducing an anti-aliasing blur filter kernel of size $n \times n$, denoted as Blurn. The Blurn filter kernel is then combined with subsampling (s) using any of the above-mentioned pooling techniques to create a Blurpooln's Anti-aliasing that can be integrated into the down-sampling operation to improve shift-equivariance and reduce the loss of information caused by aggressive filtering. However, it can also introduce blurring and reduce the image's sharpness, which can be limiting in some applications.

The main reason behind the lack of shift-invariance in Convolution Neural Networks is down-sampling [30], or stride [32]. Downsampling, as a linear operation, involves sampling image pixels at fixed grid positions and discarding the rest. As depicted in Fig.2, the disparity between down-sampling an image and its shifted counterpart can be substantial. This discrepancy arises from the fact that shifting an image can alter

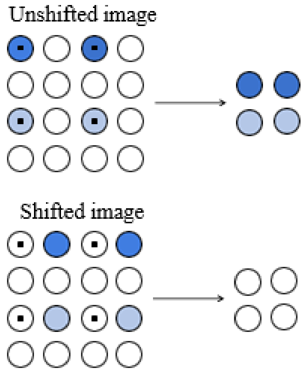

Fig. 2. Conventional linear downsampling

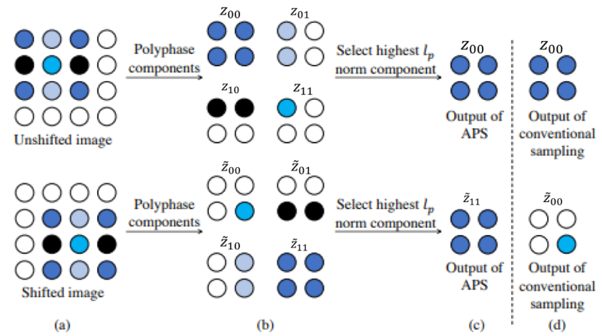

Fig. 3. APS on single channel input. (a) Image and Shifted counterpart. (b) Two images having equivalent sets of polyphase components. (c) By choosing the component with the highest lp normalization, Adaptive Polyphase sampling returns the same output for both images. (d) The output of conventional sampling in contrast.

the pixel intensities positioned over the sampling grid. Data augmentation [34] stands as an alternative method capable of enhancing a CNN's output robustness against shifts. Training the network using randomly shifted versions of input images [30], [32] can enhance its invariance, albeit limited to image patterns observed during training [30].

H. Adaptive Polyphase Pooling

Within our U-Net Architecture, we introduced Adaptive Polyphase Sampling (APS) as pooling layers. Adaptive Polyphasing Sampling (APS) is a downsampling technique proposed by Chaman and Dokamic [35] makes use of selecting an adaptive sampling grid based on some set rules for a subsampling feature map which generates a downsampled output with highest energy. The main idea of APS is predicated on a straightforward finding that a time-frequency subsampled (T-F) patch and subsampling its version that has been shifted by one bin can be varied when bins are sampled at identical fixed locations and surpasses blurring-based methods in terms of classification consistency. This mechanism underlines the absence of shift-invariance caused due to subsampling operations and implements perfect shift-invariance in CNN classifiers

1) Shift Equivariance: The shift caused in the input of a convolutional network output, which creates a similar like shift in the model's output is called as the Shift Equivariance property. Regardless of the position in frame, it is used to reconstruct objects accurately.

2) Downsampling in Adpative Polyphase Sampling: Consider a 1-D signal $s_0(n)$ with stride 2 and its discrete-time Fourier transform (DTFT). The signal can be downsampled into 4 subparts (Eq.1), denoted by phase components $\{z_{ij}\}_{(i,j)=0}^{1}$, and one pixel-shifted version denoted by $\{\tilde{z}_{ij}\}_{(i,j)=0}^{1}$ (ref Fig.3).

$$\begin{aligned} z'_{00} &= z_{11}(n_1-1, n_2-1), & z'_{10} &= z_{01}(n_1, n_2-1) \\ z'_{01} &= z_{10}(n_1-1, n_2), & z'_{11} &= z_{00}(n_1, n_2) \end{aligned} \tag{1}$$

The aim is to choose polyphase components in a permutation-invariant manner from $\{z_{ij}\}$ and $\{\tilde{z}_{ij}\}$ such that the two sets are identical but for different orderings. Let's denote the chosen polyphase component as p with $p = 2$. The Adaptive Polyphase Sampling (APS) ensures that the subsampled output for x and $\tilde{x}$ (1-pixel shifted polyphase component) is the same.

With $p = 2$, APS uses the following criterion to determine its output:

$$z_{\text{aps}} = z_i \tag{2}$$

$$\text{where } i = \arg\max_j \{\|(|z_i|)^p\}\}_{j=0}^{1}$$

The chosen polyphase components z_{ij} and $\tilde{z}_{ij}$ are permutation-invariant, and APS ensures that the subsampled outputs are identical despite different orderings.

3) Upsampling in Adaptive Polyphase Sampling: One of the issues of downsampling is that it reduces input image resolution which can be problematic in handling medical images, where image resolution is key in finding observations. To solve this issue, we make use of adaptive Polyphase upscaling which makes the creation of shift-equivariant models possible. Adaptive polyphase Upsampling is done to restore the features to their original spatial location after performing downsampling operation. If there is a downsampled feature map f obtained after performing Adaptive Polyphase downsampling from s, the output (Eq.3) of the upsampling layers o would be as follows:

$$\text{Poly}(\boldsymbol{o})_{m^\star} = \begin{cases} \boldsymbol{f}, & \text{if } m^\star = \arg\max_{m\in\{0,1\}} p_\theta(\boldsymbol{m} = m \mid \boldsymbol{s}) \\ \boldsymbol{0}, & \text{otherwise.} \end{cases} \tag{3}$$

I. Performance Metrics

In image segmentation, accuracy and Intersection over Union (IoU) are two frequently used performance metrics. Accuracy is a general metric that assesses the overall effectiveness of a segmentation model, whereas IoU is a more precise statistic that quantifies the spatial overlap between the predicted and ground truth segmentation.

1) Accuracy: The percentage of all pixels that are correctly classified is known as accuracy. It is calculated by dividing the total number of pixels in the image by the number of pixels that were correctly classified. Although accuracy is a simple and understandable metric, it has some drawbacks. For instance, by simply predicting all pixels as background, it is possible to achieve a high accuracy score.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

2) Intersection of Union: Intersection of Union(IOU) is defined as the overlap between the predicted and ground truth segmentations divided by the union of the two segmentations. The computation of this involves dividing the total number of pixels classified as foreground in either the predicted or ground truth segmentation by the number of pixels that are accurately classified as foreground.

$$IoU = \frac{TP}{TP + FP + FN} \tag{5}$$

IV. Proposed Methodology

By incorporating Adaptive Polyphase Pooling into a modified U-Net architecture, we provide a method for precise brain MRI skull-stripping that improves shift-invariance and guarantees stable performance at various image scales.

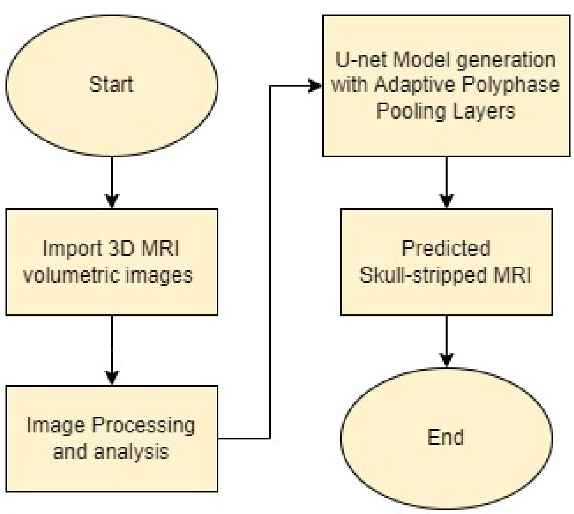

Fig. 4. Flowchart of the proposed methodology

Fig.4 represents the actual flow of our proposed system. It starts with taking 3D volumetric MRI images as input along with the labeled masks. Resizing the brain images to 256x256x256 pixels and correcting the image orientation to align with the mask's orientation and then the processed images were passed into the modified U-Net architecture with adaptive polyphase pooling layers to get the final desired output, i.e skull stripped image of the Brain MRI.

A. Dataset

The proposed model utilizes the Internet Brain Segmentation Repository (IBSR) Dataset [36], which contains T1-weighted MRI brain scans and manually segmented subcortical structures. It includes 18 MRI brain scans from healthy 20-30-year-old subjects, as well as manually segmented masks of 43 subcortical structures such as the hippocampus, amygdala, thalamus, caudate nucleus, and putamen. The IBSR Dataset is one of the most widely used datasets for brain image segmentation and analysis research, having been used to develop and evaluate new automated segmentation methods, study the effects of aging and disease on brain structure, and compare the performance of various MRI imaging techniques.

B. U-Net

The U-net architecture with skip layers breaks down a 2D slice of a brain MRI image and outputs a skull-stripped slice of a 3D image. Fig.5 depicts the suggested neural network architecture. Using encoders, features were obtained across multiple resolutions, and the MRI image was then rebuilt using decoders. Since each layer had an image size of 8x8, 16x16, 32x32, 64x64, 128x128, and 256x256, the features were extracted at several scales during the encoding phase. The channel size in the network used to extract the crucial brain features from the non-brain tissues was set to 1 for input and output MRI images. Using feature descriptors like edges, color, and texture, a supervised learning method was used to precisely identify regions of interest (ROI) to realize the segmentation process. This method aided in the recognition of intricate structures within the brain, effectively distinguishing them from non-brain tissues. The architecture incorporates adaptive polyphase sampling to enhance feature extraction and improve segmentation accuracy. The capacity of the model to capture subtle details is improved by this sampling technique, which dynamically modifies the sampling strategy across various scales. The bottleneck block area (shown by the outlined black box in Figure. 5). In order to improvise, we added more skip layers, as shown in Fig. 5 which enabled the decoder to receive both high-level and low-level characteristics from the encoder. To effectively image segmentation of images, the network's kernel size is smaller. The stride was fixed to two and the signal amplitude was maintained with zero padding. The activation function of the rectified linear unit (ReLU) was employed.

For our proposed methodology, we have taken Accuracy and IOU as performance metrics. Both accuracy and IoU are advantageous for evaluating image segmentation models. However, it is important to remember that they are not mutually exclusive. A model's accuracy score may be high but its IoU score low, or vice versa. As a result, it is crucial to consider both metrics when evaluating the performance of an image segmentation model.

V. Experimental Results

An optimal batch size of 32 is chosen, considering various variables such as dataset size, available computing power, and

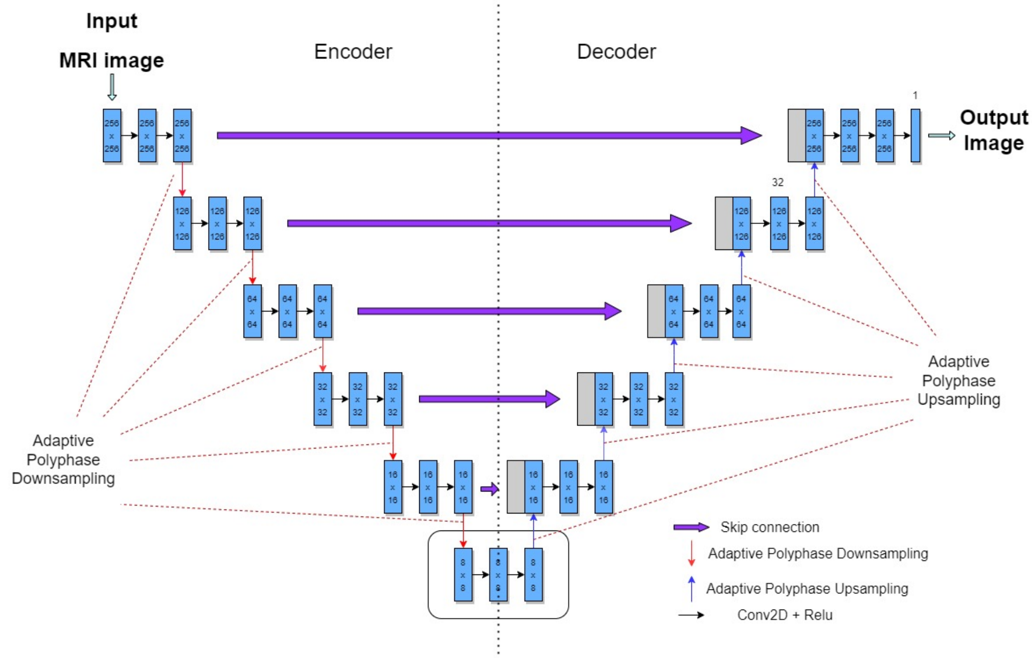

Fig. 5. Proposed U-Net architecture

U-Net model requirements.

TABLE III
LIST OF ACCURACIES ON UNET WITHOUT ADAPTIVE POOLING LAYERS

Pixel Range	0	64	128	192	256
0	99.21	98.23	98.27	98.21	98.19
64	74.32	98.45	97.13	88.23	91.22
128	96.34	92.1	98.49	99.1	93.2
192	98.23	97.56	79.34	94	98
256	94.9	98.65	95.55	96.23	94.13

TABLE IV
LIST OF ACCURACIES ON UNET WITH ADAPTIVE POOLING LAYERS

Pixel Range	0	64	128	192	256
0	99.27	99.26	99.27	99.21	99.19
64	99.35	99.45	99.2	99.37	99.37
128	99.41	99.37	99.49	99.39	99.4
192	99.1	99	99.28	99.3	99.22
256	99.33	99.39	99.46	99.39	99.35

In both Table.III and Table.IV, the values along the x-axis show displacement along the x-axis in the image in pixels, while the values along the y-axis show displacement along the y-axis in the image in pixels. Fig.6 displays a 2D slice of the side view of the original 3d volumetric MRI image without skull-stripping, the actual mask of the skull-stripped Brain MRI image, and the predicted mask by the proposed model with the Adaptive Pooling Subsampling technique The proposed model achieved an impressive 99.2% accuracy when analyzing 2D slices from a single view of 3D brain MRI images. Evaluated brain MRI images from three views: front, side, and top. The model was run three times to determine

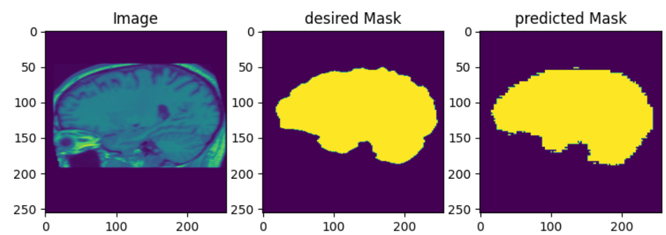

Fig. 6. displays a 2d slice of the side view of the original 3d volumetric MRI image without skull-stripping, the actual mask of the skull-stripped Brain MRI image and the predicted mask by the proposed model with the Adaptive Pooling Subsampling technique

the average result. Table.III displays a list of accuracies of the side view of the 3d brain MRI images along two axes predicted using our proposed architecture without the Adaptive pooling layers.Based on the observations from Table.III, it is evident that there are inconsistencies in predictions when the same brain MRI image is shifted along either of the two axes. This variation in accuracy when the same image is shifted leads to reduced performance and inaccurate outcomes. Inaccurate skull stripping of brain MRI images can lead to a cascade of downstream errors, including overestimation or underestimation of brain volumes, mislabeled lesions, and false tumor detections. To address this challenge, robust skull stripping methods are needed that can accurately identify and remove skull tissues while preserving the integrity of the brain parenchyma. Table.IV illustrates the list of accuracies along the same axes and for the same view predicted using our proposed architecture with Adaptive Pooling layers. It can

TABLE V
COMPARISON OF SKULL-STRIPPING METHODS/SOFTWARE

Sr. No	Method/Software Used	Time taken for Skull-Stripping	Accuracy	Parameter Efficiency
1	Deep Learning Architecture	2-3 mins	99.2%	High
2	Free Surfer	15 mins	95%	Medium
3	FSL	16 mins	93%	Medium
4	CAT	20 mins	90%	Low

be observed that our model becomes shift-invariant and gives higher consistency in the accuracy i.e. 99.2% everywhere in the view even if the brain MRI image is shifted in the same plane. The final predicted output closely resembles the actual skull-stripped mask for the brain MRI image, as depicted in Fig.6. TableV shows the method or the software used for skull-stripping of Brain MRI images, the corresponding time taken to give the final predicted output, the accuracy in accomplishing the task and the efficiency in performing the skull-stripping task. The model's excellent accuracy and dependability in skull-stripping tasks for brain MRI analysis is attributed to the combination of supervised learning, specific architectural adjustments, and shift-invariance mechanisms. One important method that has evolved to guarantee the model's correctness in 3D Brain MRI images that strip the skull is called Adaptive Polyphase Pooling (APP).

Through the introduction of shift-invariance, APP addressed the inherent shift-sensitivity of classic CNNs, allowing consistent predictions even in the case of small picture changes along two axes within the same plane. By using adaptive sample grids, this method transformed feature extraction and improved the model's capacity to accurately represent important brain structures. It made predictions that were consistent across different picture sizes, which greatly enhanced the generalization and resilience of the model. Additionally, the APP reduced downsampling difficulties while maintaining picture details that are necessary for precise skull-stripping. In the end, the incorporation of APP represented a significant breakthrough that increased the model's prediction accuracy and stability for complex brain MRI analysis.

Notably, a CNN-based approach demonstrated innovation, but further advancements could be pursued by incorporating transfer learning techniques, enhancing adaptability across diverse imaging conditions. Hybrid models that combine supervised and unsupervised learning demonstrate versatility, with potential improvements possible by incorporating attention mechanisms to refine focus during segmentation. Ensemble learning strategies [37] have been shown to be effective, but investigating active learning methods could improve training data utilization.

VI. CONCLUSION AND FUTURE WORK

As a crucial step in medical image segmentation, we tackled the significant issue of skull-stripping in 3D volumetric Brain MRI images in this study. Additionally, the approach incorporates supervised learning techniques and a modified U-Net architecture, with Adaptive Pooling layers to enhance the model's robustness and shift-invariance. When evaluated on two-dimensional brain MRI slices, our strategy outperformed other approaches in terms of accuracy. According to the findings, the suggested model analyzed 2D slices of 3D brain MRI data with an astounding accuracy of 99.2%. When the brain MRI picture was moved along the two axes in the same plane, the model continuously produced good accuracy results, according to a number of test perspectives. This shift-invariance guarantees consistent outcomes despite minute changes in the input data, which is a major accomplishment. To sum up, our study has created new opportunities for enhancing the effectiveness and calibre of medical image processing when it comes to brain MRI examinations. One of the most significant challenges was the high number of training parameters, this can be significantly reduced by incorporating wavelets. By combining wavelet transformations with the proposed approach, the model could achieve even higher accuracy and precision in the segmentation of intricate brain structures.

ACKNOWLEDGEMENT

We have considered the dataset [36] solely for training and testing the proposed model and doesn't involve any human trials.

REFERENCES

[1] Q. Wu. Research on deep learning image processing technology of second-order partial differential equations. *Neural Computing and Applications*, 35, 03 2022.
[2] G. Murtaza, L. Shuib, A. W. Abdul Wahab, G. Mujtaba, H. F. Nweke, M. A. Al-garadi, F. Zulfiqar, G. Raza, and N. A. Azmi. Deep learning-based breast cancer classification through medical imaging modalities: state of the art and research challenges. *Artificial Intelligence Review*, 53(3):1655–1720, Mar 2020.
[3] S. Perumal and V. Thambusamy. Preprocessing by contrast enhancement techniques for medical images. *International Journal of Pure and Applied Mathematics*, 118:3681–3688, 01 2018.
[4] R. C. Gonzalez and R. E. Woods. *Digital image processing*. Prentice Hall, Upper Saddle River, N.J., 2008.
[5] A. Aljuaid and M. Anwar. Survey of supervised learning for medical image processing. 3(4), may 2022.
[6] S. Shurrab and R. Duwairi. Self-supervised learning methods and applications in medical imaging analysis: a survey. *PeerJ Computer Science*, 8:e1045, 2022.
[7] S. Joshi, B. Partibane, W. A. Hatamleh, H. Tarazi, C. S. Yadav, and D. Krah. Glaucoma detection using image processing and supervised learning for classification. *Journal of Healthcare Engineering*, 2022:2988262, Mar 2022.
[8] X. Chen, X. Wang, K. Zhang, K. M. Fung, T. C. Thai, K. Moore, R. S. Mannel, H. Liu, B. Zheng, and Y. Qiu. Recent advances and clinical applications of deep learning in medical image analysis. *Medical Image Analysis*, 79:102444, 2022.
[9] W. Wöber, P. D. Tibihika, C. Olaverri Monreal, L. Mehnen, P. Sykacek, and H. Meimberg. Comparison of unsupervised learning methods for natural image processing. 07 2019.
[10] K. Yamawaki, Y. Sun, and X. H. Han. Blind image super resolution using deep unsupervised learning. *Electronics*, 10(21), 2021.
[11] M. Moreno López, J. M. Frederick, and J. Ventura. Evaluation of mri denoising methods using unsupervised learning. *Frontiers in Artificial Intelligence*, 4, 2021.
[12] P. Schmid. Segmentation of digitized dermatoscopic images by two-dimensional color clustering. *IEEE Transactions on Medical Imaging*, 18(2):164–171, 1999.

[13] National Library of Medicine - National Institutes of Health. National library of medicine - national institutes of health, n.d. Retrieved October 15, 2023.

[14] C. Fennema-Notestine, I. B. Ozyurt, C. P. Clark, S. Morris, A. Bischoff-Grethe, M. W. Bondi, T. L. Jernigan, B. Fischl, F. Segonne, D. W. Shattuck, R. M. Leahy, D. E. Rex, A. W. Toga, K. H. Zou, and G. G. Brown. Quantitative evaluation of automated skull-stripping methods applied to contemporary and legacy images: effects of diagnosis, bias correction, and slice location. *Hum Brain Mapp*, 27(2):99–113.

[15] A. Klein, S. S. Ghosh, B. Avants, B. T. Yeo, B. Fischl, B. Ardekani, J. C. Gee, J. J. Mann, and R. V. Parsey. Evaluation of volume-based and surface-based brain image registration methods. *Neuroimage*, 51(1):214–20, May 15 2010.

[16] N. F. Kalkers, N. Ameziane, J. C. Bot, A. Minneboo, C. H. Polman, and F. Barkhof. Longitudinal brain volume measurement in multiple sclerosis: rate of brain atrophy is independent of the disease subtype. *Arch Neurol*, 59(10):1572–6, Oct 2002.

[17] M. Wels, Y. Zheng, M. Huber, J. Hornegger, and D. Comaniciu. A discriminative model-constrained em approach to 3d mri brain tissue classification and intensity non-uniformity correction. *Phys Med Biol*, 56(11):3269–300, Jun 7 2011.

[18] P. M. Thompson, M. S. Mega, R. P. Woods, C. I. Zoumalan, C. J. Lindshield, R. E. Blanton, J. Moussai, C. J. Holmes, J. L. Cummings, and A. W. Toga. Cortical change in alzheimer's disease detected with a disease-specific population-based brain atlas. *Cerebral Cortex*, 11(1):1–16, 01 2001.

[19] L. Zhao, U. Ruotsalainen, J. Hirvonen, J. Hietala, and J. Tohka. Automatic cerebral and cerebellar hemisphere segmentation in 3d mri: adaptive disconnection algorithm. *Med Image Anal*, 14(3):360–72, Jun 2010.

[20] R. Zivadinov, F. Bagnato, D. Nasuelli, S. Bastianello, A. Bratina, L. Locatelli, K. Watts, L. Finamore, A. Grop, M. Dwyer, M. Catalan, A. Clemenzi, E. Millefiorini, R. Bakshi, and M. Zorzon. Short-term brain atrophy changes in relapsing-remitting multiple sclerosis. *J Neurol Sci*, 223(2):185–93, Aug 30 2004.

[21] H. Rusinek, M. J. de Leon, A. E. George, L. A. Stylopoulos, R. Chandra, G. Smith, T. Rand, M. Mourino, and H. Kowalski. Alzheimer disease: measuring loss of cerebral gray matter with mr imaging. *Radiology*, 178(1):109–14, Jan 1991.

[22] P. Tanskanen, J. M. Veijola, U. K. Piippo, M. Haapea, J. A. Miettunen, J. Pyhtinen, E. T. Bullmore, P. B. Jones, and M. K. Isohanni. Hippocampus and amygdala volumes in schizophrenia and other psychoses in the northern finland 1966 birth cohort. *Schizophr Res*, 75(2-3):283–94, Jun 15 2005.

[23] Karuppanagounder Somasundaram and P. Kalavathi. Skull stripping from mri of head scans based on 2 d region growing. 2011.

[24] C. Balaji and S. Veni. Automatic skull stripping from mri of human brain using deep learning framework for the diagnosis of brain related diseases. *International Journal of Intelligent Systems and Applications in Engineering*, 11(4):439–445, Sep. 2023.

[25] Y. Gu and K. Li. A transfer model based on supervised multi-layer dictionary learning for brain tumor mri image recognition. *Frontiers in Neuroscience*, 15, 2021.

[26] X. He, A. Q. Wang, and M. R. Sabuncu. Neural pre-processing: A learning framework for end-to-end brain mri pre-processing. pages 258–267, 2023.

[27] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1(4):541–551, 1989.

[28] Y. L. Cun, B. Boser, J. S. Denker, R. E. Howard, W. Habbard, L. D. Jackel, and D. Henderson. *Handwritten Digit Recognition with a Back-Propagation Network*, page 396–404. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1990.

[29] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86:2278 – 2324, 12 1998.

[30] A Azulay and Y Weiss. Why do deep convolutional networks generalize so poorly to small image transformations?, 2019.

[31] O Kayhan and J Gemert. On translation invariance in cnns: Convolutional layers can exploit absolute spatial location, 2020.

[32] R Zhang. Making convolutional networks shift-invariant again, 2019.

[33] Dominik Scherer, Andreas Müller, and Sven Behnke. Evaluation of pooling operations in convolutional architectures for object recognition. In Konstantinos Diamantaras, Wlodek Duch, and Lazaros S. Iliadis, editors, *Artificial Neural Networks – ICANN 2010*, pages 92–101, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg.

[34] R Simonyan and A Zisserman. Very deep convolutional networks for large-scale image recognition, 2015.

[35] A Oppenheim, J Buck, and R Schafer. *Discrete-Time Signal Processing*, volume 2. Hall, P, Upper Saddle River, NJ, 2001.

[36] Ibsr: Tool/resource info. NITRC. Accessed 1 November 2023.

[37] Vatsala Anand, Sheifali Gupta, Deepali Gupta, Yonis Gulzar, Qin Xin, Sapna Juneja, Asadullah Shah, and Asadullah Shaikh. Weighted average ensemble deep learning model for stratification of brain tumor in mri images. *Diagnostics*, 13(7), 2023.