

Attention Res-UNet: Attention Residual UNet With Focal Tversky Loss for Skin Lesion Segmentation

Aasia Rehman, University of Kashmir, India*

 <https://orcid.org/0000-0003-2065-5699>

Muheet A. Butt, University of Kashmir, India

Majid Zaman, University of Kashmir, India

ABSTRACT

During a dermoscopy examination, accurate and automatic skin lesion detection and segmentation can assist medical experts in resecting problematic areas and decrease the risk of deaths due to skin cancer. In order to develop fully automated deep learning model for skin lesion segmentation, the authors design a model Attention Res-UNet by incorporating residual connections, squeeze and excite units, atrous spatial pyramid pooling, and attention gates in basic UNet architecture. This model uses focal tversky loss function to achieve better trade off among recall and precision when training on smaller size lesions while improving the overall outcome of the proposed model. The results of experiments have demonstrated that this design, when evaluated on publicly available ISIC 2018 skin lesion segmentation dataset, outperforms the existing standard methods with a Dice score of 89.14% and IoU of 81.16%; and achieves better trade off among precision and recall. The authors have also performed statistical test of this model with other standard methods and evaluated that this model is statistically significant.

KEYWORDS

Attention gates, Dermoscopy, Lesion Segmentation, Medical Images, Residual connections, Squeeze and Excitation

1. INTRODUCTION

Automated medical image segmentation has been extensively studied in the medical image analysis field since radiologists usually have to manually look for malignancy in a pool of images and match cancer-related features to candidate tumors (Hesamian et al., 2019b). In this process of diagnosis, some features may get easily missed in many cases and there also exist some important features that cannot be extracted visually but can be with an automated system. The disagreement among different radiologists in the segmentation of images is reported to be between 2% to 49% (Hesamian et al., 2019b). All these facts are the primary motivation for designing a computer aided segmentation

DOI: 10.4018/IJDSST.315756

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

model for segmenting medical images accurately. As per 2022 statistics related to cancer article (Siegel et al., 2022), approximately 1,918,030 cancer cases will be detected of which 99,780 will be of melanoma skin cancer. A report by (Siegel et al., 2021), the death rate of cancer has dropped from its peak in 1991 to 2018 because of a reduction in smoking and the development of models that lead to early detection and medication of cancer. If detected in the early stages the survival rate for melanoma skin cancer is 93% according to 2021 statistics (Siegel et al., 2021). The unconstrained development of abnormal cells leads to skin cancer which can further spread to the rest of the organs of the body. Cancer of the skin is often defined as fatal melanoma or benignant (Wei et al., 2019). Among various kinds of skin cancers, melanoma cancer is of utmost deadly, accounting for a large percentage of skin cancer deaths (Siegel & Miller, 2019). Because of the fatal nature of melanoma, it has attracted ample research and clinical attention.

Photography, dermoscopy, confocal scanning laser microscopy (CSLM), optical coherence tomography (OCT), ultrasonography, magnetic resonance imaging (MRI), and spectroscopic imaging are presently utilized to help dermatologists in skin lesion identification (Hasan et al., 2020). Dermatologists commonly visually scrutinize the produced photos using the specified procedures to detect cancerous skin, which is typically thought to be a laborious and time consuming task. Dermoscopy, which has been in use for over 20 years, has increased the diagnostic rate when compared to viewable surveillance solely (Mayer, 1997). The ABCD benchmark assists non-professionals in distinguishing benignant melanocytic naevi from melanoma while screening skin lesions (Abbasi et al., 2004). End-to-end computerized technologies that can precisely segment skin lesions of all sorts are very appropriate to emulate the clinical ABCD benchmark. Computer aided diagnostic programs have been designed to support medical experts and enhance accuracy. In many diagnostic centers, computer aided diagnosis has become the habitual clinical practice for diagnosing abnormal growth of lesions in medical images. Computer aided technologies for dermoscopic medical images are typically most of the time made of more than one unit including image acquisition, image pretreatment, segmenting images, feature mining, and classification units (Fan et al., 2017)(Jalalian et al., 2017). The precise segmentation of lesions in Dermoscopy images from the normal skin serves a significant role in gaining unique and exemplary characteristics of melanoma areas of interest in Dermoscopy images (Wei et al., 2019). Several skin lesion segmentation models (Yuan, 2017)(Ebenezer & Rajapakse, 2018) (Berseeth, 2017)(Bi et al., 2017) that utilize deeper neural networks with convolutional layers have been developed and have substantially improved the outcome of segmentation. Though, automatically segmenting skin lesions from their background is still considered an open research problem or a challenging problem because of the following reasons:

- Deep CNNs can't be effectively trained due to a lack of data in the medical imaging field. Deep learning methods have greatly enhanced the performance of medical image segmentation however these networks usually require a huge amount of labelled input samples to execute the training. Gathering such huge quantity of labelled dataset in medical image analyses is oftentimes a vigorous task and annotating new images will also be tedious and costly.
- There is an imbalance in labeling Dermoscopy images as some of the lesions occupy a small portion of the images and cannot be discriminated from the background so the recall rate of models must be improved. A low recall means small size skin lesions are missed.
- Skin tumors generally occur in different colors, shapes, and locations, have uneven and vague boundaries, and have a low distinctive feature with the neighboring skin. Also the existence of several artifacts like body hair, frames, air bubbles, blood vessels, shadows, haphazard lighting, markers, and ink. (Hasan et al., 2020) can escalate the complexity of segmentation models.

To the best of our understanding, these issues have yet to be fully resolved. In this research paper, the authors focus on how to deal with data imbalance and boost the performance of computerized segmentation of skin lesions. The authors propose an ensemble Attention Res-UNet model by

combining a variant of UNet (Ronneberger et al., n.d.) known as ResUNet (Z. Zhang et al., 2018) which has demonstrated state-of-art outcomes for segmenting of road images. The authors use ResUNet (Z. Zhang et al., 2018) as a basis for our architecture and incorporate squeeze and excitation units (Hu et al., 2020), Atrous Spatial Pyramid Pooling (ASPP) (He et al., 2014), Attention gates as in (Oktay et al., 2018) in ResUNet (Z. Zhang et al., 2018) along with focal Tversky loss function (Abraham & Khan, 2019) to segment small size skin lesions effectively. The authors evaluate their model on the publically available ISIC 2018 skin lesion dataset (Codella et al., 2018) (Tschandl et al., 2018) as it has a data imbalance issue with huge variations in skin lesion sizes. In ISIC 2018 dataset, the lesions occupy between 4.84% and 5.43% of the Dermoscopy images (Abraham & Khan, 2019). The results of implementation have demonstrated that our ensemble design outperforms the baseline model known as UNet and various other state-of-art approaches for ISIC 2018 skin lesion dataset (Codella et al., 2018) (Tschandl et al., 2018). In this paper, the authors assess the efficacy of their design using Dice score, Intersection over Union (IoU), recall, and precision.

The remaining paper is structured as follows: 2nd part explains the literature work done in the area of segmentation of skin lesions, 3rd part describes the proposed model in detail and the Focal Tversky Loss function, part 4th provides the implementation outcomes obtained by our designed model and compares the obtained results with other existing standard methods, part 5th concludes the paper.

2. LITERATURE REVIEW

This part reviews various standard deep-learning networks designed for segmenting skin lesions. Although in the early years several numbers of models were designed to tackle the problem of skin lesion segmentation. Four major categories of traditional models include region expansion (Celebi et al., 2008), thresholding (Yüksel & Borlu, 2009), effective-contour-based (Mete & Sirakov, 2010), and clustering (B. S. Lin et al., 2018) based models. However, these methods depend significantly on hand-engineered features, necessitating subject expertise and expert contributions (Wei et al., 2019). As a result, traditional segmentation algorithms still have limited discriminating capabilities and give a poor performance (Wei et al., 2019). Deep Learning models have gained a lot of research interest in medical image analysis including lung nodule segmentation (C. Zhang et al., 2019), segmentation of the brain (Jiang et al., 2019), and brain tumor classification (Sultan et al., 2019). In 2015 popularly known deep learning architecture in biomedical image segmentation was designed by O. Ronneberger et al. known as UNet (Ronneberger et al., n.d.). UNet a variant of FCN is an encoder-decoder architecture that demonstrates good experimental results despite smaller size datasets.

Machine Learning and deep learning have already been utilized for various healthcare tasks like in (Altaf et al., 2021) (Ashraf et al., 2021) (Fayaz et al., 2022) (Riyaz et al., 2022) (Fayaz et al., 2021) (Altaf et al., 2022) (Rehman et al., 2022). In general, several deep learning models have been developed for the segmentation of different medical image modalities which have been discussed in our review (Rehman et al., 2021). Several deep learning models in the domain of detecting and segmenting lesions on the skin have been designed by researchers. B. S. Lin et al. (B. S. Lin et al., 2018) utilized ISBI 2017 data to evaluate and analyze two lesion segmenting techniques: UNet, based on histogram equalization and c-means clustering. After evaluation and analyzing both methods UNet did better than the clustering approach. Y. Yuan in (Yuan, 2017) designed a model based on a deep full convolution-deconvolution neural network called CDNN (Shelhamer et al., 2017) for mechanized segmentation of skin lesions. The proposed model focused on building appropriate network architecture so that the proposed deep-learning network can address dermoscopic images of diverse acquiring situations rather than utilizing complicated preprocessing and post-processing methods and hand-engineered features (Yuan, 2017). Yuan, Yading, et al. (Yuan et al., 2017) developed a 19-level deep CNN-based end-to-end entirely mechanized approach for segmenting skin lesions. To solve the problem of asymmetry among foreground and background pixels, they created a new loss function based on Jaccard Distance. They trained with ISBI 2016 using 5-fold cross-validation and analyzed their results using ISBI 2016

and PH2 datasets. On these two datasets, the implementation outcome verified that the developed technique outperformed existing models. However, they claimed that the method performed poorly in several difficult cases such as images with low contrast. Md. Zahangir et al. (Alom et al., 2018) developed a UNet architecture based on recurrent neural networks and recurrent residual networks namely RU-Net and R2U-Net accordingly. The designed models are assessed using three different standard datasets namely retinal dataset, skin lesion, and pulmonary nodule datasets. The results have shown that the designed models perform exceptionally well on different segmentation datasets when related to other UNet and residual based networks. N. Abraham et al. (Abraham & Khan, 2019) developed an enhanced attention UNet by integrating an image pyramid to retain contextual information. They also proposed a general Focal tversky loss function to overcome the trouble of data imbalance in medical image segmentation (Abraham & Khan, 2019) which showed good trade off among precision and recall when the model is trained on smaller size lesions. Experiments were done on the breast cancer dataset and skin lesion (Codella et al., 2018)(Tschandl et al., 2018) dataset in which lesions take up only a smaller image area. Results obtained from experiments demonstrated that the developed method enhanced segmentation accuracy when related to base model UNet with 25.7 and 3.6% for the breast cancer dataset and 2018 skin lesion (ISIC) datasets correspondingly (Abraham & Khan, 2019). Azad, Reza, et al. (Azad et al., 2019) Drafted an extended edition of UNet known as Bi-directional ConvLSTM UNet with densely linked convolutions, for segmenting medical images. They have utilized the pros of UNet and used bi-directional ConvLSTM (BConvLSTM) in the skip connections to aggregate feature maps that are drawn from the complementary encoder pathway and from the previous decoder layer in a non-linear way (Azad et al., 2019). The authors also employed dense convolutions in the encoder's last level to improve feature transmission and encourage feature reuse and employed batch normalization for accelerating the convergence speed of the proposed model. The designed architecture was assessed using three distinct datasets: retinal, skin lesion, and lung nodule datasets, and it outperformed the state of art results. Xie et al. (Xie et al., 2020) designed the MB-DCNN model based on associating segmentation and classification of lesions which can mutually propagate semantic features and granular lesion masks among mask facilitated classification model (mask-CN) and a granular segmentation model to enhance the accuracy of the segmentation task. Ibtehaz, N. et al (Ibtehaz & Rahman, 2020) proposed some alterations to the base model UNet and developed a novel model known as MultiResUNet. They evaluated their model on five different datasets and compared their results with the base model UNet. The proposed model demonstrated better performance than the basic model UNet. X. Guo et al. (Guo et al., 2020) designed a novel model known as a complementary model with adaptive receptive field learning to deal with the hole and shrink problems of previous techniques for the segmentation of lesions. The authors developed a forefront design to identify malignant lesions and a back-ground design to conceal non-melanoma areas, rather than approaching the segmentation problem separately. The authors proposed adaptive atrous convolution and knowledge aggregation modules to deal with hole and shrinkage issues. A unique mutual loss was presented that takes advantage of the interdependence amongst the foreground and background models, allowing them to impact each other reciprocally. The designed model was assessed on ISIC 2018 skin lesion dataset (Codella et al., 2018)(Tschandl et al., 2018) and experiments have demonstrated good dice of 86.4%. The designed network outperformed the existing standard techniques by a large margin. Jha et al. (Jha et al., 2020) devised a new model named as DoubleUNet formed by the combination of two UNet models stacked on one another. First UNet used a pre-trained VGG-Net model on the ImageNet dataset which could be easily transferred to another task. Second UNet was embedded to capture more semantic features effectively. Kumar et al. (Kumar et al., 2020) devised a new deep learning based segmentation model for segmenting skin lesions by making use of illumination invariance of various tissues to enhance accuracy. Tomar et al. (Tomar et al., 2021) developed an attention based feedback model known as FANet that combines preceding iterations mask with features of currently executing iteration. Preceding iterations mask is then utilized to produce hard attention to the features obtained at various levels. The designed model also permits correcting

the estimated output repetitively through the testing phase. Wu et al. (Wu et al., 2022) developed architecture based on the classical UNet model known as the FAT-Net network which combined an additional transformer unit to efficiently encapsulate hard dependencies and global context features. Moreover, the proposed architecture also utilized the memory efficient synthesis part as well as a feature adaption subsystem to improve feature blending among the neighboring features by invoking efficacious channels and preventing the unimportant back-ground features.

The development of methods to cope with class imbalance is a dominant topic of study in the segmentation of medical images. To reduce class imbalance, the focus loss as in (T. Y. Lin et al., 2020) prevents a large number of simple negative samples from diminishing the gradient. Due to the smaller regions of interest present in medical images, however, it has difficulties balancing accuracy and recall in practice. More discriminative approaches, just as Attention gated networks (Oktay et al., 2018), have been proposed in research attempts to solve smaller ROI segmentation (Abraham & Khan, 2019). CNNs with attention gates (AGs) may be trained end to end, emphasizing the marked region concerning the classification aim. These gates produce soft region suggestions at test time in order to emphasize important ROI characteristics while suppressing feature activation by irrelevant regions. During testing these attention gates develop soft area recommendations to emphasize the important area of interest features and withhold activation features by uninteresting regions (Abraham & Khan, 2019). To overcome the problem of class asymmetry, small ROI segmentation, and to enhance performance the authors associate residual attention UNet with Squeeze-and-Excitation block (Hu et al., 2020) and attention gates (Oktay et al., 2018) using the focal Tversky loss function (Abraham & Khan, 2019).

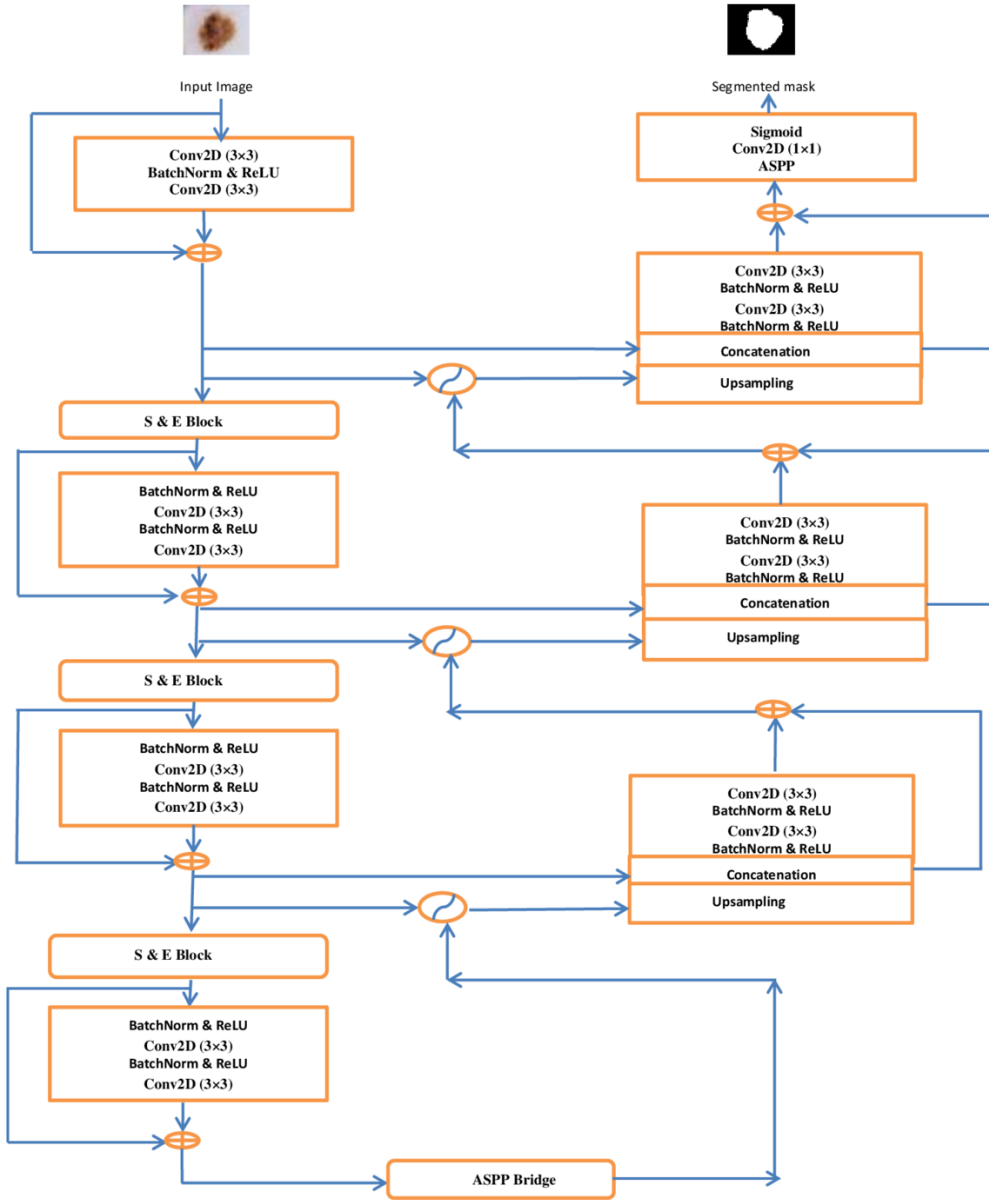
3. PROPOSED METHODOLOGY

This section of the paper explains the suggested approach for segmenting skin lesions:

3.1 Model Architecture

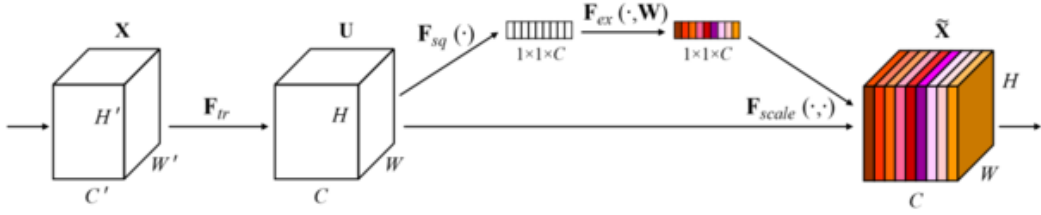
The proposed model design is built on the deep residual network ResUNet (Zhang et al., 2018) that makes use of Squeeze-and-Excitation block (Hu et al., 2020), atrous spatial pyramid pooling, attention gates, and Focal Tversky loss function for segmenting skin lesions. The classical network for biomedical semantic segmentation, UNet (Ronneberger et al., n.d.) is made up of two parts analysis and synthesis. The CNN structure is followed throughout the analysis part for feature extraction. In the synthesis part also referred to as the expansion route, an upsampling layer is succeeded by each deconvolutional layer (Hesamian et al., 2019a). The most significant feature of UNet is shortcut connections among equal-resolution layers in the analysis and synthesis paths (Hesamian et al., 2019a). As a result of these skip connections, high resolution features are transferred to deconvolutional layers (Hesamian et al., 2019a). The information flow across different layers is propagated using residual units as in ResUNet (Z. Zhang et al., 2018) which allows for building deeper models resolving the vanishing gradient problem in the analysis part. This increases the interdependencies across channels while lowering the computational cost. The designed model is comprised of one basic unit followed by three blocks in the analysis part, atrous spatial pyramid pooling used as a bridge between the two parts, and three blocks in the synthesis part. Figure 1 depicts the proposed model in the block diagram form. From the block diagram of the model, it can be interpreted that each block in the analysis path comprises a pair of convolutional units including a batch-norm layer, a Rectified linear unit, and a layer of the convolution operation. In addition to this, the input and output of each analysis unit are connected by identity mapping. The Squeeze-and-Excitation block (Hu et al., 2020) takes the output of each block in the analysis part of the model. Then atrous spatial pyramid pooling (ASPP) (He et al., 2014)(Chen et al., 2018) is used as a bridge between the two parts of the model which enables the filter's field of vision to be expanded to encompass wider domains.

Figure 1. Block Diagram of proposed Attention Res-UNet. Here  represents attention gate



The synthesis part of the model consists of similar blocks. Attention gates are applied prior to each residual unit in the synthesis phase to increase the effectiveness of feature maps which is followed by upsampling of low level feature maps and concatenation of feature maps from the corresponding analysis block. The results of the last block of the synthesis path are sent to ASPP and then the segmentation map is generated using 1x1 convolution along with the sigmoid activation function. Different parts of the proposed model are briefly explained below:

Figure 2. Squeeze and Excitation Block as in original paper (Hu et al., 2020)



Residual Blocks: Researchers have suggested that training deeper networks can increase model performance. However, it is often seen that training deeper models is extremely challenging and can lead to problems like vanishing gradients (Z. Zhang et al., 2018). To tackle this problem, He et al. (He et al., 2016) developed a deep residual learning strategy, which makes use of identity mapping to make training deeper models easier. Olaf Ronneberger et al. (Ronneberger et al., n.d.) designed a UNet structure in which features from various levels are aggregated to enhance the performance of segmentation. UNet is the basic model used for biomedical image segmentation as it concatenated low level features with high level features. Zhang et al (Zhang et al., 2018) designed a deeper residual U-Net network called ResUNet that combines the best features of both the deep residual learning framework and UNet (Ronneberger et al., n.d.). The inclusion of residual blocks makes training of deeper networks simpler, and the skip connection in the model assists in information flow without damaging the architecture of the neural network, allowing it to function much better on the semantic segmentation job while decreasing the parameters (Z. Zhang et al., 2018). These are the reasons for using ResUNet as the underlying framework for our network.

Squeeze and Excitation Block (Hu et al., 2020): Squeeze-and-Excitation block, a type of self-attention mechanism that was developed by J. Hu et al (Hu et al., 2020) is based on the channel diameters. CNN's can concentrate more on aggregated relevant context with long-range relationships or semantically significant areas due to self-attention methods. Medical image analysis systems may become more reliable by leveraging attention to concentrate on key clinical feature areas. SE also known as the channel recalibration technique (Rao et al., 2021), which means the amplitude of each channel may be deliberately adjusted based on the values of other channels. The Squeeze-and-Excitation (Hu et al., 2020) block's objective is to guarantee that the model's sensitivity to essential features increases while irrelevant features are suppressed. As depicted in Figure 2, the first step is squeezing which uses global average pooling (GAP) to collect global data in each channel. The concept of "global" refers to the reduction of spatial coordinates (altitude and width) and the usage of just one pooled value per channel. Excitation (active calibration) is the second step, which uses a globally pooled feature matrix and two subsequent fully-connected layers with one ReLU layer in between. The end channel size is similar to the input channel size, whereas the transitional channel width is 1/16 of the input channel width. After this sigmoid activation is applied to the last feature vector it is then multiplied back to the original full feature map (Rao et al., 2021). The steps of the Squeeze-and-Excitation block (Hu et al., 2020) are shown below:

$$\begin{aligned}
 f_{ch} &= GAP(F) \\
 a_{ch} &= \sigma(FC_{c/r \rightarrow c}(ReLU(FC_{c \rightarrow c/r}(f_{ch})))) \\
 SA(F) &= F * a_{ch}
 \end{aligned}$$

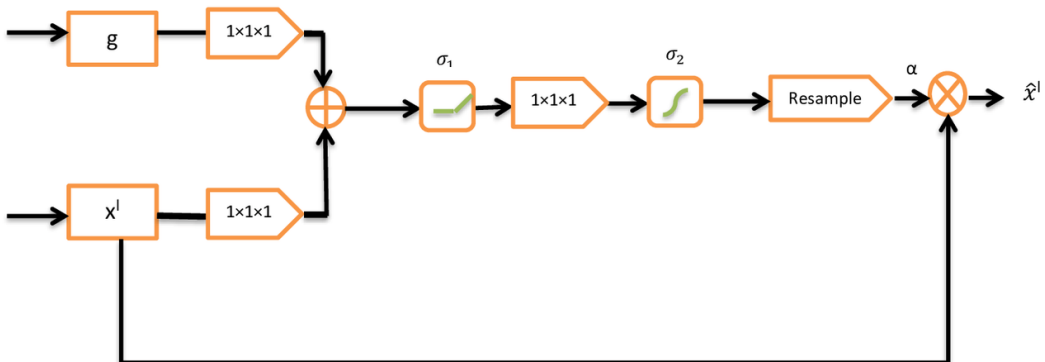
Equations are taken from (Rao et al., 2021), here σ implies sigmoid activation, GAP stands for global average pooling function, r for intermediate channel reduction ratio, and FC for fully-connected layers with input or output channels defined. Here $F \in \mathbb{R}^{C \times H \times W}$, $f_{ch} \in \mathbb{R}^{C \times 1 \times 1}$, and $a_{ch} \in \mathbb{R}^{C \times 1 \times 1}$. The result of the final multiplication is spread over the spatial dimensions. In the proposed model the authors have utilized SE block along with residual units intending to enhance the network's performance by increasing effective generalization across diverse datasets.

ASPP Block (Chen et al., 2018): The concept of ASPP (Chen et al., 2018) is based on the success of spatial pyramidal pooling (Sultan et al., 2019), which was used to re-sample features at various scales. The contextual information is collected at multiple scales in ASPP (Chen et al., 2018), and the input feature map is fused using numerous parallel atrous convolutions with varying rates (Chen et al., 2018). By offering multi-scale information, the ASPP approach has demonstrated remarkable results on several segmentation tasks. As a result, for the semantic segmentation task in our model, the authors employ ASPP to capture the important multi-scale information. ASPP is employed as a bridge between the analysis and synthesis path of our model as shown in Figure 1.

Attention Gate (Oktay et al., 2018): Attention gates are most frequently utilized in Natural Language Processing, image analysis, and knowledge graphs (Oktay et al., 2018). Trainable Attention is classified as soft and hard attention (Oktay et al., 2018). Hard attention (Mnih et al., 2014) such as cyclic region classifier and pruning is frequently not differentiable and generally relies on reinforcement learning to modify parameters, making network training more challenging. Soft-Attention (Oktay et al., 2018), on the other hand, is probabilistic and uses normal back-propagation instead of Monte Carlo sampling. Additive soft attention, for example, is employed in sentence-sentence rephrasing (Oktay et al., 2018) and, more recently, picture categorization (Oktay et al., 2018). The network has the fullest possible feature representation at the deepest stage of the analysis path. However, spatial information in high-level output maps tends to be lost when cascaded convolutions and non-linearities are used (Abraham & Khan, 2019). As a result, it is hard to minimize false detections for smaller lesions with large variations in size and shape. To tackle this problem our model utilizes soft attention gates as shown in Figure 3 to recognize spatial features from lower level feature maps and pass them on to the synthesis path.

For every pixel i , Attention gates (Oktay et al., 2018) generate coefficients $\alpha_i \in [0, 1]$ which scales input feature mappings x_i^l at layer l to yield high level semantic features $\hat{x}_i^l$ (Abraham & Khan, 2019). The schematic diagram of the soft attention gate used by (Abraham & Khan, 2019) in our

Figure 3. Schematic diagram of additive attention gate (AG) adapted from (Oktay et al., 2018). To transmit important attributes to the synthesis layer output $\hat{x}_i^l$ input features x_i^l are scaled with attention coefficients α_i . Bilinear interpolation is used to calculate feature map resampling



model is shown in Figure 3. To identify important regions a gating signal g is applied to every pixel i . The additive attention is represented mathematically as follows:

$$q_{attn}^l = \varphi^T(\sigma_1(W_x^T x_i^l + W_g^T g_i + b_g)) + b_\varphi \quad (1)$$

$$\alpha_i^l = \sigma_2(q_{attn}^l(x_i^l, g_i; \Theta_{attn})) \quad (2)$$

Here $\sigma_2(x_{i,c}) = 1 \div (1 + \exp(-x_{i,c}))$ is identified as sigmoid activation. The attention gate is recognized by a collection of parameters Q_{attn} including linear transformations $W_x \in R^{F_l \times F_{int}}$, $W_g \in R^{F_g \times F_{int}}$, $\varphi \in R^{F_{int} \times 1}$ and bias terms $b_\varphi \in R$, $b_g \in R^{F_{int}}$ (Oktay et al., 2018). Attention coefficients which are linear q_{attn}^l , are calculated by element by element addition and $1 \times 1 \times 1$ linear convolutions for the input tensors (Abraham & Khan, 2019). The attention coefficient α_i uses element wise multiplication to scale the lower lever feature maps and keep only significant activation maps (Abraham & Khan, 2019). Along the synthesis path, these pruned features are aggregated with up-sampled output maps at each scale (Abraham & Khan, 2019). Finally, the authors would like to point out that attention gate parameters may be trained using conventional back-propagation updates (Oktay et al., 2018), which eliminates the requirement for sampling-based update techniques like those employed in hard-attention (Yüksel & Borlu, 2009).

3.2 Focal Tversky Loss Function (FTL)

In the field of medical image analysis, there are a few things to keep in mind. For medical image segmentation, the dice score (DSC) is the most often employed assessment measure. The overlap among the expected mask and the ground-truth is measured by the Dice coefficient. The lowest overlap among the anticipated value and the ground reality is referred to as dice loss (Abraham & Khan, 2019). Among various flaws of dice loss, the most important is that it weights both false positive (FP) and false negative (FN) analysis equally. This in practice yields segmentation maps with higher precision but low recall. To enhance the recall rate, false negatives must be weighted higher than false positives when the data is extremely unbalanced and the area of interest is smaller, such as skin lesions. The Tversky similarity index is a Dice coefficient generalization that provides for more versatility in balancing false positives and false negatives (Abraham & Khan, 2019). One more difficulty with the Dice Loss is that it has trouble segmenting smaller areas of interest because it doesn't dedicate much time towards loss. To tackle the above mentioned issue the authors utilize the loss function as proposed in (Abraham & Khan, 2019) known as the Focal Tversky loss function (FTL), parameterized using γ and may be used to adjust the contrast between easy background and hard area of interest training instances (Abraham & Khan, 2019). The authors in (Abraham & Khan, 2019) have expressed FTL as:

$$FTL_c = \sum_c (1 - TI_c)^{1/\gamma} \quad (3)$$

Here γ can be in the range $[1, 3]$ and TI_c is the Tversky index which is expressed as follows:

Table 1. Results of our proposed model and performance comparison with other standard models on ISIC 2018 (Codella et al., 2018)(Tschandl et al., 2018) skin lesion dataset

Model	Dice Score (%)	IoU (%)	Recall	Precision
UNet (Ronneberger et al., n.d.)	67.4	54.9	70.8	-----
R2UNet(Alom et al., 2018)(Azad et al., 2019)	67.9	58.1	79.2	-----
Attention-R2UNet (Alom et al., 2018)(Azad et al., 2019)	69.1	59.2	72.6	-----
SegNet (Badrinarayanan et al., 2017)	70.92	-----	80.56	-----
UNet++ (Zhou et al., 2018)	80.88	73.19	84.50	86.48
Attention U-Net (Oktay et al., 2018)	82.05	73.46	85.16	86.45
BCDU (d=1) (Azad et al., 2019)	84.70	-----	78.30	-----
ResUNet (Z. Zhang et al., 2018) + FTL	84.502±0.471	75.818±0.56	89.24±0.86	84.87±1.21
BCDU (d=3) (Azad et al., 2019)	85.1	-----	78.5	-----
Attention U-Net + Multi-Input + FTL (Abraham & Khan, 2019)	85.6	-----	89.7	85.8
Our Model (Without Data Augmentation)	87.11± 0.49	79.46± 0.64	89.88± 0.79	88.50± 0.84
Our Model (With Data Augmentation)	89.14	81.16	91.60	88.86

$$TI_c = \frac{\sum_{i=1}^N p_{ic} g_{ic} + \epsilon}{\sum_{i=1}^N p_{ic} g_{ic} + \alpha \sum_{i=1}^N p_{i\bar{c}} g_{ic} + \beta \sum_{i=1}^N p_{ic} g_{i\bar{c}} + \epsilon} \quad (4)$$

In equation (4), the p_{ic} represents the likelihood that pixel i belongs to the lesion class namely c , and the $p_{i\bar{c}}$ denotes the likelihood that pixel i belongs to non-lesion class $\bar{c}$. Similarly, g_{ic} & $g_{i\bar{c}}$ denotes the actual label of pixel i that belongs to tumor class c and non-lesion class $\bar{c}$ respectively.

In effect, the Focal Tversky loss is unaffected if a pixel is miss-labeled using a large value of the Tversky index (TI). The focal Tversky loss will fall considerably if the Tversky index is low and the pixel is miss-labeled. The authors in (Abraham & Khan, 2019) have observed that when the value γ is 4/3 it shows the best results. Therefore, the authors trained their model with the same value of $\gamma = 4/3$. In the experiments, the authors use α and β as 0.7 and 0.3 accordingly as used in (Abraham & Khan, 2019). If $\alpha = \beta = 0.5$ then tversky index TI is same as dice coefficient and if $\gamma = 1$ then Focal Tversky loss is same as tversky loss (Abraham & Khan, 2019).

4. EXPERIMENTS AND RESULTS

The authors evaluate the Attention-ResUNet model by training, validating, and testing the model utilizing a publically available skin lesion dataset. Then the results of the proposed model are compared with existing standard models.

4.1 Experiments

Our model is evaluated using ISIC 2018 dataset taken from skin lesion analysis toward melanoma detection grand challenge dataset (Codella et al., 2018),(Tschandl et al., 2018). The assembled collection includes 2594 pictures of various forms of skin lesions, all of which have been expertly annotated. The dataset consists of images of various resolutions with an average resolution of 2166×3188 pixels (Abraham & Khan, 2019). However, the images have been rescaled to 192×256 pixels retaining the average aspect ratio for computation. Train test split is used to divide the dataset into training, validation, and testing sets, with 75% going to training and validation and the remaining 25% going to testing. The authors have also used 5-fold cross validation for reporting Dice Score, IoU, precision, and recall. Our model is trained for 150 epochs with starting learning rate of 0.01, a batch of size 16, and a threshold of 0.5. Stochastic gradient descent is used as an optimizer with decay 10^{-6} and the Focal tversky loss function (FTL)(Abraham & Khan, 2019) is used as a loss. All the parameters are optimized by manually training the models with multiple sets of hyper parameters and assessing their outcomes. Keras with Tensor-Flow backend is utilized as the programming framework for our model.

Data Augmentation: To further boost the accuracy of the designed architecture the authors have applied data augmentation to our training dataset. The Data augmentation technique is often utilized to thwart the problem of data inadequacy in the medical image analysis domain. Data augmentation increases the number of training samples during training. On the training dataset, the authors have used several augmentation techniques such as center crop, rotation, transpose, scaling, and elastic transform. Each training image was converted into 15 different images making a total of 21784 images for training.

4.2 Results

Table 1 depicts the achievement of trained models and it is evident from the table that the baseline model known as UNet (Ronneberger et al., n.d.), for segmentation of skin lesions performs below average. Table 1 demonstrates the results of UNet (Ronneberger et al., n.d.), R2UNet (Alom et al., 2018), Attention-R2UNet (Alom et al., 2018), BCDU(Azad et al., 2019),SegNet (Badrinarayanan et al., 2017), UNet++ (Zhou et al., 2018), Attention UNet (Oktay et al., 2018), Attention UNet with FTL (Abraham & Khan, 2019), ResUNet (Z. Zhang et al., 2018) and our proposed model Attention Res-UNet. The authors have assessed the model performance using Dice Score, IoU, recall, and precision for ISIC 2018 skin lesion dataset.

Figure 4 demonstrates a few of the output segmentation masks of our model for the ISIC 2018 dataset. In Figure 4, second and third rows are the examples of tiny lesion segmentation for which the model achieves dice scores of 0.89419 and 0.97349 respectively. The results (dice scores) for tiny lesions as shown in second and third rows of Figure 4 demonstrate the effectiveness of our model, that it segments small size lesions efficiently. Figure 5 (a) and 5 (b) shows the Dice Score vs. Epoch and IoU vs. Epoch models of our Attention Res-UNet architecture for ISIC 2018 skin lesion segmentation dataset. The designed architecture named as Attention Res-UNet converges faster (after 50 epochs) as can be seen from Figure 5(a) and 5(b). In Figure 5(a) and 5(b), it can be seen that the dice score and IoU on validation data varies during the training phase. This is due to the fact that the validation data has few images that are completely distinct from images present in training dataset thus the deep network in the initial iterations of training has some difficulties for segmenting images. Table 1 demonstrates that our designed model achieves higher values for Dice score, IoU, recall, and precision. The proposed network has out-performed the baseline model UNet by a significant margin with Dice score of 89.14%, $87.11\% \pm 0.49$, and IoU of 81.16%, $79.46\% \pm 0.644$ with and without data augmentation respectively. Thus our model enhances overall performance for lesion segmentation in the skin dataset and acquired a fair balance between recall and precision.

For analysis of variance, the authors have conducted an ANOVA test and reported the p-value in order to illustrate the statistical significance of our experimental results. The ANOVA test is carried

Figure 4. Some results of the proposed model for the ISIC 2018 dataset

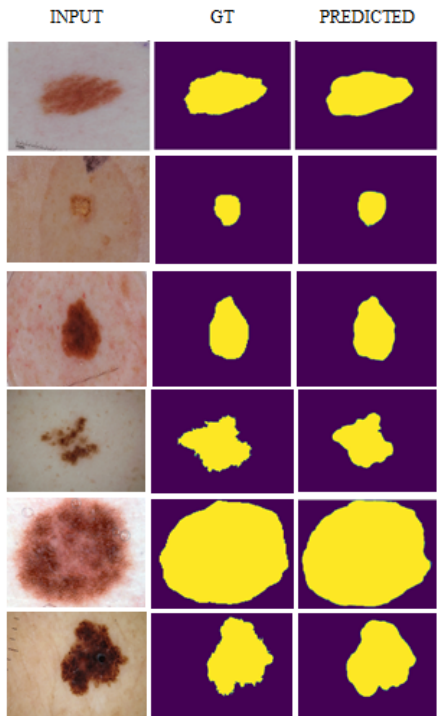
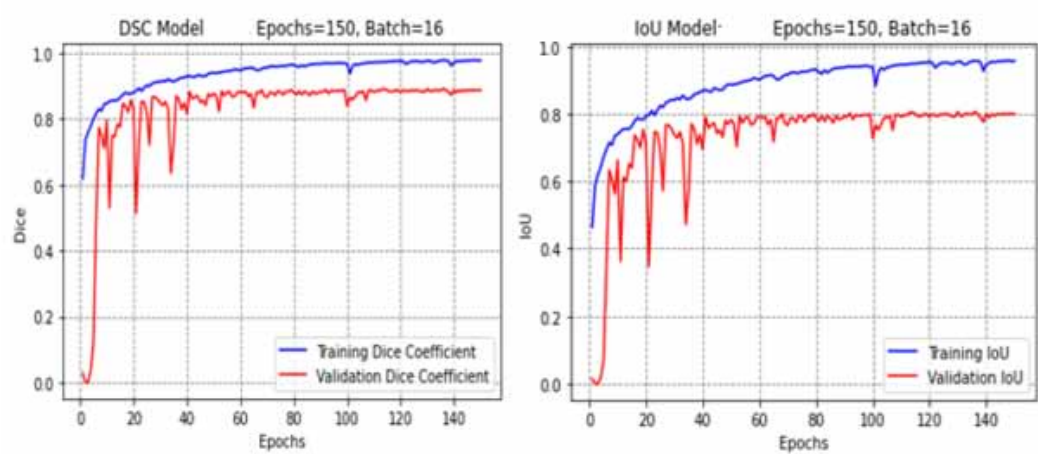


Figure 5. Dice score and IoU models of Attention Res-UNet for ISIC 2018 dataset



out between the dice scores of our proposed model (with and without data augmentation) and the other baseline deep learning techniques. The p-values of all the methods against our designed model are shown in Table 2 which demonstrates that all p-values are less than 0.05 where the authors dismiss the null hypotheses and take the alternate hypotheses which state that there is a significant difference among our method and other baseline methods. Hence this proves the statistical significance of our model experimental results.

Table 2. P-values of baseline models concerning our model without and with Data Augmentation

Model type	p-value(without Data Augmented model)	p-value(with data augmented model)
UNet (Ronneberger et al., n.d.)+FTL	1.54165E-05	3.56E-08
R2UNet(Alom et al., 2018)(Azad et al., 2019)	1.29E-39	1.17E-51
Attention-R2UNet (Alom et al., 2018)(Azad et al., 2019)	7.23E-43	7.75E-57
BCDU (d=1) (Azad et al., 2019)	1.06014E-05	1.64E-07
ResUNet (Z. Zhang et al., 2018) + FTL	2.44E-05	1.01E-07
BCDU (d=3) (Azad et al., 2019)	0.024593	1.09E-05
Attention U-Net + Multi-Input + FTL (Abraham & Khan, 2019)	0.024703	0.000797

5. CONCLUSION

The authors proposed a novel Attention Res-UNet in this paper, which is a network designed to meet the requirement for more precise segmentation of skin lesions in Dermoscopy images. Attention Res-UNet utilizes the best features of residual blocks, Squeeze-and-Excite block, spatial pyramid pooling, and Attention gates. The proposed network in experimental evaluation using ISIC 2018 skin lesion dataset outperforms the other standard networks based on UNet and Residual networks in terms of generating semantically correct predictions and representing a fair balance between precision and recall. The network achieved a dice score of 89.14% and IoU of 81.16% with an improvement of 21.74% and 26.26% in dice score and IoU respectively when equated with the classical model UNet. The designed model has also been evaluated utilizing the ANOVA test which calculated the p-values of dice scores among our model and other models. The observed p-values are lower than 0.05 which proves the statistical significance of our model. The developed network can be a solid basis for additional studies to build a therapeutically effective technique to be used for other types of medical imaging datasets.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

FUNDING STATEMENT

The authors did not receive support from any organization for the submitted work

REFERENCES

- Abbasi, N. R., Shaw, H. M., Rigel, D. S., Friedman, R. J., McCarthy, W. H., Osman, I., Kopf, A. W., & Polsky, D. (2004). Early diagnosis of cutaneous melanoma: Revisiting the ABCD criteria. *Journal of the American Medical Association*, 292(22), 2771–2776. doi:10.1001/jama.292.22.2771 PMID:15585738
- Abraham, N., & Khan, N. M. (2019). A novel focal tversky loss function with improved attention u-net for lesion segmentation. *Proceedings - International Symposium on Biomedical Imaging*, (pp. 683–687). IEEE. doi:10.1109/ISBI.2019.8759329
- AlomM. Z.HasanM. YakopcicC.TahaT. M.AsariV. K. (2018). *Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation*. <https://arxiv.org/abs/1802.06955>
- Altaf, I., Butt, M. A., & Zaman, M. (2022). ETL for disease indicators using brute force rule-based NLP algorithm and metadata exploration. *International Journal of Advanced Technology and Engineering Exploration*, 9(90).
- Altaf, I., Butt, M. A., & Zaman, M. (2021). A Pragmatic Comparison of Supervised Machine Learning Classifiers for Disease Diagnosis. *2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)*, 1515–1520. IEEE. doi:10.1109/ICIRCA51532.2021.9544582
- Ashraf, M., Ahmad, S. M., Ganai, N. A., Shah, R. A., Zaman, M., Khan, S. A., & Shah, A. A. (2021). Prediction of Cardiovascular Disease Through Cutting-Edge Deep Learning Technologies: An Empirical Study Based on TENSORFLOW, PYTORCH and KERAS. In D. Gupta, A. Khanna, S. Bhattacharyya, A. E. Hassanien, S. Anand, & A. Jaiswal (Eds.), *International Conference on Innovative Computing and Communications* (pp. 239–255). Springer. doi:10.1007/978-981-15-5113-0_18
- Azad, R., Asadi-Aghbolaghi, M., Fathy, M., & Escalera, S. (2019). Bi-directional ConvLSTM U-net with densely connected convolutions. *Proceedings - 2019 International Conference on Computer Vision Workshop, ICCVW 2019*, 406–415. IEEE. doi:10.1109/ICCVW.2019.00052
- Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12), 2481–2495. doi:10.1109/TPAMI.2016.2644615 PMID:28060704
- BersethM. (2017). *Skin Lesion Analysis Towards Melanoma Detection*. 1–4. <https://arxiv.org/abs/1703.00523>
- BiL.KimJ.AhnE.FengD. (2017). *Automatic Skin Lesion Analysis using Large-scale Dermoscopy Images and Deep Residual Networks*. 6–9. <https://arxiv.org/abs/1703.04197>
- Celebi, M. E., Kingravi, H. A., Iyatomi, H., Aslandogan, Y. A., Stoecker, W. V., Moss, R. H., Malters, J. M., Grichnik, J. M., Marghoob, A. A., Rabinovitz, H. S., & Menzies, S. W. (2008). Border detection in dermoscopy images using statistical region merging. *Skin Research and Technology: Official Journal of International Society for Bioengineering and the Skin (ISBS) [and] International Society for Digital Imaging of Skin (ISDIS) [and] [ISSI]. International Society for Skin Imaging*, 14(3), 347–353. doi:10.1111/j.1600-0846.2008.00301.x PMID:19159382
- Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2018). DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4), 834–848. doi:10.1109/TPAMI.2017.2699184 PMID:28463186
- Codella, N. C. F., Gutman, D., Celebi, M. E., Helba, B., Marchetti, M. A., Dusza, S. W., Kalloo, A., Liopyris, K., Mishra, N., Kittler, H., & Halpern, A. (2018). Skin lesion analysis toward melanoma detection: A challenge at the 2017 International symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC). *Proceedings - International Symposium on Biomedical Imaging, 2018-April*, 168–172. IEEE. doi:10.1109/ISBI.2018.8363547
- EbenezerJ. P.RajapakseJ. C. (2018). *Automatic segmentation of skin lesions using deep learning*. 1–4. <https://arxiv.org/abs/1807.04893>
- Fan, H., Xie, F., Li, Y., Jiang, Z., & Liu, J. (2017). Automatic segmentation of dermoscopy images using saliency combined with Otsu threshold. *Computers in Biology and Medicine*, 85, 75–85. doi:10.1016/j.compbmed.2017.03.025 PMID:28460258

- Fayaz, S. A., Zaman, M., & Butt, M. A. (2021). To Ameliorate Classification Accuracy using Ensemble Distributed Decision Tree (DDT) Vote Approach: An Empirical discourse of Geographical Data Mining. *Procedia Computer Science*, 184, 935–940. doi:10.1016/j.procs.2021.03.116
- Fayaz, S. A., Zaman, M., Kaul, S., & Butt, M. A. (2022). Is Deep Learning on Tabular Data Enough? An Assessment. *International Journal of Advanced Computer Science and Applications*, 13(4), 466–473. doi:10.14569/IJACSA.2022.0130454
- Guo, X., Chen, Z., & Yuan, Y. (2020). Complementary Network with Adaptive Receptive Fields for Melanoma Segmentation. *Proceedings - International Symposium on Biomedical Imaging*. IEEE. doi:10.1109/ISBI45749.2020.9098417
- Hasan, M. K., Dahal, L., Samarakoon, P. N., Tushar, F. I., & Martí, R. (2020). DSNet: Automatic dermoscopic skin lesion segmentation. *Computers in Biology and Medicine*, 120(April), 103738. doi:10.1016/j.combiomed.2020.103738 PMID:32421644
- He, K., Zhang, X., Ren, S., & Sun, J. (2014). Spatial pyramid pooling in deep convolutional networks for visual recognition. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8691 LNCS(PART 3), 346–361. IEEE. doi:10.1007/978-3-319-10578-9_23
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016-Decem*, 770–778. IEEE. doi:10.1109/CVPR.2016.90
- Hesamian, M. H., Jia, W., He, X., & Kennedy, P. (2019a). Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges. *Journal of Digital Imaging*, 32(4), 582–596. doi:10.1007/s10278-019-00227-x PMID:31144149
- Hesamian, M. H., Jia, W., He, X., & Kennedy, P. J. (2019b). Atrous Convolution for Binary Semantic Segmentation of Lung Nodule. *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1015–1019. IEEE. doi:10.1109/ICASSP.2019.8682220
- Hu, J., Shen, L., Albanie, S., Sun, G., & Wu, E. (2020). Squeeze-and-Excitation Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(8), 2011–2023. doi:10.1109/TPAMI.2019.2913372 PMID:31034408
- Ibtehaz, N., & Rahman, M. S. (2020). MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation. *Neural Networks*, 121, 74–87. doi:10.1016/j.neunet.2019.08.025 PMID:31536901
- Jalalian, A., Mashohor, S., Mahmud, R., Karasfi, B., Ramli, A. R. B., Engineering, C. S., Putra, U., Science, H., Putra, U., & Branch, Q. (2017). Review article : Foundation And Methodologies In Computer-Aided. *EXCLI Journal*, 113–137.
- Jha, D., Riegler, M. A., Johansen, D., Halvorsen, P., & Johansen, H. D. (2020). DoubleU-Net: A deep convolutional neural network for medical image segmentation. *Proceedings - IEEE Symposium on Computer-Based Medical Systems, 2020-July*, 558–564. IEEE. doi:10.1109/CBMS49503.2020.00111
- Jiang, J., Hu, Y. C., Liu, C. J., Halpenny, D., Hellmann, M. D., Deasy, J. O., Mageras, G., & Veeraraghavan, H. (2019). Multiple resolution residually connected feature streams for automatic lung tumor segmentation from CT images. *IEEE Transactions on Medical Imaging*, 38(1), 134–144. doi:10.1109/TMI.2018.2857800 PMID:30040632
- Kumar, A., Hamarneh, G., & Drew, M. S. (2020). Illumination-based transformations improve skin lesion segmentation in dermoscopic images. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 3132–3141. IEEE. doi:10.1109/CVPRW50498.2020.00372
- Lin, B. S., Michael, K., Kalra, S., & Tizhoosh, H. R. (2018). Skin lesion segmentation: U-Nets versus clustering. *Symposium Series on Computational Intelligence, SSCI 2017 - Proceedings*, , 1–7. IEEE. doi:10.1109/SSCI.2017.8280804
- Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2020). Focal Loss for Dense Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2), 318–327. doi:10.1109/TPAMI.2018.2858826 PMID:30040631

- Mayer, J. (1997). Systematic review of the diagnostic accuracy of dermatoscopy in detecting malignant melanoma. *The Medical Journal of Australia*, 167(4), 206–210. doi:10.5694/j.1326-5377.1997.tb138847.x PMID:9293268
- Mete, M., & Sirakov, N. M. (2010). Lesion detection in demoscopy images with novel density-based and active contour approaches. *BMC Bioinformatics*, 11(S6, SUPPL. 6), S23. doi:10.1186/1471-2105-11-S6-S23 PMID:20946607
- Mnih, V., Heess, N., Graves, A., & Kavukcuoglu, K. (2014). Recurrent Models of Visual Attention. *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, 2204–2212. IEEE.
- OktayO.SchlemperJ.Le FolgocL.LeeM.HeinrichM.MisawaK.MoriK.McDonaghS.HammerlaN. Y.KainzB. GlockerB.RueckertD. (2018). *Attention U-Net: Learning Where to Look for the Pancreas. Midl.* https://arxiv.org/abs/1804.03999
- Rao, A., Park, J., Woo, S., Lee, J. Y., & Aalami, O. (2021). Studying the Effects of Self-Attention for Medical Image Analysis. *Proceedings of the IEEE International Conference on Computer Vision*, 3409–3418. IEEE. doi:10.1109/ICCVW54120.2021.00381
- Rehman, A., Ahmed Butt, M., & Zaman, M. (2021). A Survey of Medical Image Analysis Using Deep Learning Approaches. *5th International Conference on Computing Methodologies and Communication (ICCMC)*, 1334–1342. IEEE. doi:10.1109/ICCMC51019.2021.9418385
- Rehman, A., Butt, M. A., & Zaman, M. (2022). ChA-DenseUNet: Channel Attention DenseUNet for Polyp Segmentation. *6th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 1594–1599. IEEE. doi:10.1109/ICICCS53718.2022.9788206
- Riyaz, L., Butt, M. A., Zaman, M., & Ayob, O. (2022). Heart Disease Prediction Using Machine Learning Techniques: A Quantitative Review. In A. Khanna, D. Gupta, S. Bhattacharyya, A. E. Hassanien, S. Anand, & A. Jaiswal (Eds.), *International Conference on Innovative Computing and Communications* (pp. 81–94). Springer. doi:10.1007/978-981-16-3071-2_8
- Ronneberger, O., Fischer, P., & Brox, T. (n.d.). *U-Net: Convolutional Networks for Biomedical Image Segmentation*. 1–8.
- Shelhamer, E., Long, J., & Darrell, T. (2017). Fully Convolutional Networks for Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4), 640–651. doi:10.1109/TPAMI.2016.2572683 PMID:27244717
- Siegel, R. L., Miller, K. D., & Jemal, A. (2019). Cancer statistics, 2019. *Cancer Statistics*, 69(1), 7–34. doi:10.3322/caac.21551 PMID:30620402
- Siegel, R. L., Miller, K. D., Fuchs, H. E., & Jemal, A. (2021). Cancer Statistics, 2021. *CA: a Cancer Journal for Clinicians*, 71(1), 7–33. doi:10.3322/caac.21654 PMID:33433946
- Siegel, R. L., Miller, K. D., Fuchs, H. E., & Jemal, A. (2022). Cancer statistics, 2022. *CA: a Cancer Journal for Clinicians*, 72(1), 7–33. doi:10.3322/caac.21708 PMID:35020204
- Sultan, H. H., Salem, N. M., & Al-Atabany, W. (2019). Multi-Classification of Brain Tumor Images Using Deep Neural Network. *IEEE Access: Practical Innovations, Open Solutions*, 7, 69215–69225. doi:10.1109/ACCESS.2019.2919122
- TomarN. K.JhaD.RieglerM. A.JohansenH. D.JohansenD.RittscherJ.HalvorsenP.AliS. (2021). *FANet: A Feedback Attention Network for Improved Biomedical Image Segmentation. XX(Xx)*, 1–13. https://arxiv.org/abs/2103.17235
- Tschandl, P., Rosendahl, C., & Kittler, H. (2018). Data descriptor: The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data*, 5(1), 1–9. doi:10.1038/sdata.2018.161 PMID:30106392
- Wei, Z., Song, H., Chen, L., Li, Q., & Han, G. (2019). Attention-based denseunet network with adversarial training for skin lesion segmentation. *IEEE Access: Practical Innovations, Open Solutions*, 7, 136616–136629. doi:10.1109/ACCESS.2019.2940794
- Wu, H., Chen, S., Chen, G., Wang, W., Lei, B., & Wen, Z. (2022). FAT-Net: Feature adaptive transformers for automated skin lesion segmentation. *Medical Image Analysis*, 76, 102327. doi:10.1016/j.media.2021.102327 PMID:34923250

- Xie, Y., Zhang, J., Xia, Y., & Shen, C. (2020). A Mutual Bootstrapping Model for Automated Skin Lesion Segmentation and Classification. *IEEE Transactions on Medical Imaging*, 39(7), 2482–2493. doi:10.1109/TMI.2020.2972964 PMID:32070946
- Yuan, Y. (2017). *Automatic skin lesion segmentation with fully convolutional-deconvolutional networks*, 2–5. IEEE. 10.1109/JBHI.2017.2787487
- Yuan, Y., Chao, M., & Lo, Y.-C. (2017). Automatic Skin Lesion Segmentation Using Deep Fully Convolutional Networks With Jaccard Distance. *IEEE Transactions on Medical Imaging*, 36(9), 1876–1886. doi:10.1109/TMI.2017.2695227 PMID:28436853
- Yüksel, M. E., & Borlu, M. (2009). Accurate Segmentation of Dermoscopic Images by Image Thresholding Based on Type-2 Fuzzy Logic. *IEEE Transactions on Fuzzy Systems*, 17(4), 976–982. doi:10.1109/TFUZZ.2009.2018300
- Zhang, C., Song, Y., Liu, S., Lill, S., Wang, C., Tang, Z., You, Y., Gao, Y., Klistorner, A., Barnett, M., & Cai, W. (2019). MS-GAN: GAN-Based Semantic Segmentation of Multiple Sclerosis Lesions in Brain Magnetic Resonance Imaging. *International Conference on Digital Image Computing: Techniques and Applications, DICTA 2018*, 1–8. IEEE. doi:10.1109/DICTA.2018.8615771
- Zhang, Z., Liu, Q., & Wang, Y. (2018). Road Extraction by Deep Residual U-Net. *IEEE Geoscience and Remote Sensing Letters*, 15(5), 749–753. doi:10.1109/LGRS.2018.2802944
- Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11045 LNCS, 3–11. Springer. doi:10.1007/978-3-030-00889-5_1

Aasia Rehman is a Research Scholar in the P.G Department of Computer Sciences at Kashmir University. She has received her MTech in Computer Science and Engineering from Shri Mata Vaishno Devi University, Katra, Jammu, and her BTech in Computer Science and Engineering from Kashmir University, Srinagar. Her areas of expertise include Deep Learning, Medical Image Segmentation

Muheet Ahmed Butt is a Scientist "D" in the Post Graduate Department of Computer Science, University of Kashmir, Srinagar, India. He received his PhD degree from the University of Kashmir, J&K, India and M.Tech in Communications and Information Technology from the National Institute of Technology [NIT] Srinagar. Additionally, he holds a Bachelor of Science in Computer Science Engineering from Bangalore University, India. Email: ermuheet@gmail.com

Majid Zaman is Scientist "E" in the Directorate of Information Technology & Support System, University of Kashmir, J&K, India. He holds a PhD in Computer Science from the University of Kashmir, Srinagar, J&K, and an M.S. in Software Systems from the Birla Institute of Technology and Science (BITS), Pilani, India. In addition, he received a Bachelor's in Computer Science Engineering from BAMU, Mumbai, India. Email: zamanmajid@gmail.com