Enhancing Medical Image Segmentation using Deep Learning: Exploring State-of-the-art Models and Loss Functions for Class Imbalanced Datasets

Chevi Parsai[1], Palak Kakani[2], Shreya Vyas[3]

[1]UG Student, Department of Computer Science & Engineering, Shri Govindram Seksaria Institute of Technology and Science (SGSITS), Indore, M.P., India.
[2] UG Student, Department of Information Technology & Engineering, Shri Govindram Seksaria Institute of Technology and Science(SGSITS), Indore, M.P., India.
[3] UG Student, Department of Computer Science & Engineering, Institute of Engineering and Technology (IET DAVV), Indore, M.P., India

---***---

Abstract - *Medical image segmentation involves the extraction of specific objects from medical images, whether they are in 2D or 3D. This is done by training the model on publicly available datasets such as that of brain tumour, kidney tumour, skin lesion etc. One of the significant challenges in this field is dealing with class imbalance, particularly when tumours occupy a significantly smaller portion of the image compared to the surrounding organs or tissues. Several state of the art models have been proposed for medical image segmentation such as U-Net, Attention U-Net Efficient Net, ResUNet, SegNet, VNet etc. U-Net is a convolutional neural network (CNN) architecture consisting of encoder and decoder blocks. The prevailing loss functions such as cross-entropy loss are commonly used but lack robustness against class imbalance. Loss functions such as focal loss, combo loss are comparatively more robust to class imbalance.*

In this study, multiple Segmentation Models such as U-Net, ResUnet, Attention U-Net architecture are employed to examine their effectiveness when combined with various loss functions.

The research is conducted using two distinct medical imaging datasets: CVC Clinic DB and ISIC 2018 Skin Lesion Dataset. The datasets are trained using different architectures and their performance has been compared. Given inherent class imbalance in these datasets, loss functions such as focal loss, dice loss are employed which are more robust to class imbalance as compared to traditional loss functions such as binary cross entropy loss. The research project meticulously compares the performance of several models trained with different loss functions.

Key Words: Deep Learning, Convolutional Neural Network (CNN), Image Segmentation, Computer Vision, Machine Learning, Medical image Segmentation

1. INTRODUCTION

The fundamental concept behind image segmentation is to divide an image into distinct regions based on pixel characteristics, enabling the identification of areas of significance. Image segmentation finds applications in various domains, including facial recognition, autonomous driving, and medical imaging. Medical Image Segmentation is mainly employed to identify tumours in a colonoscopy scan/ MRI scan/ CT scan etc. Among the architectures designed for image segmentation, U-Net stands out as a widely utilized option. U-Net represents an evolved form of Convolutional Neural Network (CNN) that takes the shape of an encoder-decoder network. Leading-edge models in image segmentation encompass diverse U-Net variations like V-Net, Attention U-Net, and 3D U-Net. In the context of this project, a U-Net, Attention U-Net and ResUNet architectures are adopted, incorporating diverse loss functions. Attention enhances the model's ability to focus on relevant features and improve its segmentation performance. The objective is to assess their efficacy using two medical imaging datasets: CVC Clinic DB, and ISIC 2018 Skin Lesion Dataset. Notably, these datasets exhibit class imbalance.

While cross-entropy loss enjoys widespread adoption, its susceptibility to class imbalance is evident. Dice loss, Combo Loss and Focal Loss have emerged as alternative loss functions within the realm of medical image segmentation. Addressing the core challenge of class imbalance in medical image segmentation, particularly concerning tumours that occupy a minimal portion of scans compared to surrounding organs, necessitates a robust loss function and the architecture that learns the required features from the image. The research project entails a comprehensive comparison of the performance of different architectures against different loss functions.

1.1 Motivation

The motivation behind this research stems from the requirement to create a machine learning model which can accurately distinguish tumour from surrounding tissue in a scan. The choice of the architecture along with the choice of loss function combined affect the model performance.

Hence it is required to select the best possible combination of architecture and loss function. This research aims to contribute to the development of intelligent systems that can transform the medical field.

2. LITERATURE REVIEW

2.1 Image Classification and Machine Learning:

Image classification is a challenging task that involves teaching machine learning algorithms to recognize specific classes or objects within images. Deep learning, a subset of machine learning, plays a pivotal role in image classification. Convolutional Neural Networks (CNNs) are the cornerstone of deep learning for image classification, featuring layers for data input, hidden layers for data transformation, and output layers for predictions. CNNs are widely used for various image-related tasks, including image segmentation.

2.2 U-Net Architecture:

The U-Net architecture is a significant advancement in CNN design, tailored specifically for image segmentation. It features an encoder-decoder structure, where the encoder extracts features, and the decoder performs pixel-wise classification. The contracting path of U-Net involves convolutional layers, ReLU activation functions, and max-pooling, while the expansive path includes up-sampling and feature concatenation. U-Net's ability to seamlessly merge feature extraction and pixel classification makes it highly effective in medical image segmentation.

2.3 ResUNet Architecture:

ResUNet is an innovative architecture that combines the strengths of U-Net and Deep Residual Learning. It replaces standard convolution layers in U-Net with pre-activated residual blocks, which allows for the creation of deeper networks without the issues of vanishing or exploding gradients. ResUNet also employs skip connections for improved gradient flow during training.

2.4 Attention U-Net Architecture:

Attention mechanisms play a crucial role in computer vision, especially in image segmentation. Attention U-Net introduces a spatial attention mechanism known as the Attention Gate (AG) to focus on target structures within medical images. AG helps models suppress irrelevant image regions while emphasizing salient features, eliminating the need for explicit localization modules. Integrating Attention U-Net into segmentation architectures enhances sensitivity and accuracy.

2.5 Datasets:

Two class-imbalanced datasets, CVC Clinic DB and ISIC 2018 Skin Lesion Dataset, have been used in this research. CVC Clinic DB contains colorectal polyp images, while ISIC 2018 Skin Lesion Dataset comprises skin lesion images. Class imbalance poses a challenge in accurately segmenting tumours from surrounding tissues in these datasets.

3. BACKGROUND

3.1 Understanding Image Segmentation

Image segmentation serves as a pivotal technique within the field of computer vision, playing a crucial role in various applications such as object recognition, medical imaging, and autonomous driving. This process involves the intricate division of an image into multiple distinct segments or regions, each attributed to a specific category based on discernible criteria like colour, texture, or intensity.

In the digital landscape, every individual pixel of an image is meticulously assigned a label that corresponds to its designated segment or region. This collective effort culminates in the formation of coherent regions or structures that collectively encompass the entire visual composition. This strategy offers a finer granularity of understanding compared to conventional image classification, as it not only identifies the presence of objects but also intricately outlines their precise boundaries.

Delving deeper into the realm of image analysis, we encounter the concept of semantic segmentation, which emerges as a refined iteration of image segmentation. In this advanced process, the primary objective revolves around the pixel-level classification of an image, effectively delineating each pixel into its respective class. An intriguing facet of semantic segmentation lies in its treatment of multiple instances of the same object class as a cohesive unit. Consequently, pixels associated with the same class are portrayed with consistent colors, fostering a coherent visual representation.

Segmentation finds substantial application in diverse sectors. In medical imaging, it enables the identification and differentiation of various structures within scans, aiding in accurate diagnosis. In the context of self-driving vehicles, this technique empowers the vehicle's perception system to discern pedestrians, vehicles, and road markings, thereby ensuring safe navigation.

In summary, image segmentation, particularly the nuanced arena of semantic segmentation, emerges as a pivotal tool in extracting intricate insights from visual data. Through the meticulous allocation of labels and the creation of meaningful segments, this technique opens avenues for enhanced comprehension, paving the way for advancements across a spectrum of industries.

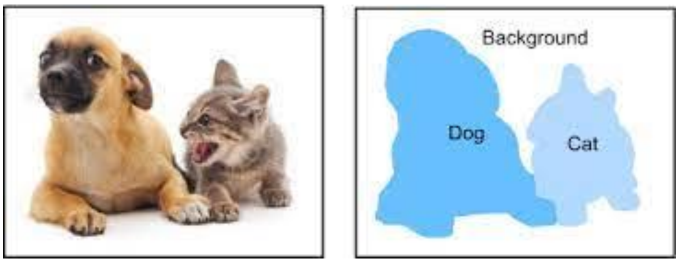


Fig-1: Image Segmentation

3.1.1 Types of Image Segmentation

Semantic Segmentation:

Semantic segmentation is a pixel-level classification technique that assigns a label to each pixel in an image based on the category or class it belongs to. In the context of medical image segmentation, semantic segmentation allows us to identify and classify different structures or regions within an image. For instance, in skin lesion analysis, it helps distinguish between the lesion itself and the surrounding healthy skin or background. Semantic segmentation provides valuable insights by creating a dense pixel-wise map of the different regions of interest.

Instance Segmentation:

Instance segmentation, on the other hand, takes image segmentation to a more granular level by not only classifying each pixel but also distinguishing between individual instances of the same class. In the context of medical imaging, instance segmentation can be particularly useful when multiple instances of a medical condition or object need to be identified separately. For example, in the case of skin lesion analysis, instance segmentation would not only classify each lesion but also assign a unique identifier to each distinct lesion within the same image.

In our research, both semantic and instance segmentation techniques were explored and evaluated, considering the specific requirements of the ISIC 2018 Skin Lesion Dataset. Semantic segmentation aided in categorizing regions of interest, such as lesions and healthy skin, while instance segmentation allowed for the precise identification of individual lesions within an image. These segmentation approaches, when combined with advanced neural network architectures and tailored loss functions, contributed to the robustness and accuracy of our skin lesion segmentation models. By leveraging the strengths of both semantic and instance segmentation, we aimed to create a comprehensive and clinically valuable solution for skin cancer screening and diagnosis.

These segmentation methods are critical components of our research methodology, enabling us to extract meaningful information from medical images and further our understanding of skin lesion analysis and diagnosis.

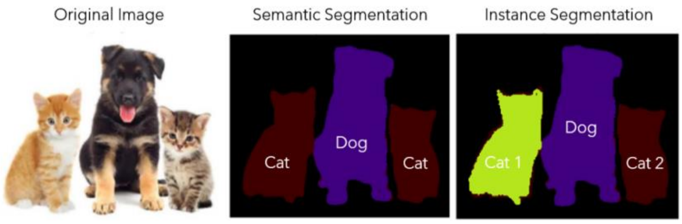


Fig-2: Types of Image Segmentation

3.2 Unveiling the U-Net Architecture

The U-Net architecture stands as a transformative evolution of the conventional convolutional neural network, ingeniously reimagined into an encoder-decoder structure, profoundly impacting the realms of feature extraction and pixel classification. This innovative design comprises two distinct segments: the contracting path and the expansive path, each equipped with encoder and decoder blocks.

Within the contracting path, the architecture commences with a pair of convolutional layers, each characterized by a 3x3 kernel size, fostering feature extraction. Following these convolutional layers, a rectified linear unit (ReLU) activation function imbues non-linearity, subsequently paving the way for a 2x2 max pooling operation. This operation, executed with a stride of 2, orchestrates an effective down-sampling process. An intriguing aspect here is the doubling of feature count with every down-sampling stride, a mechanism that amplifies the model's grasp of salient features.

Transitioning to the expansive path, a sequence of steps ushers in up-sampling through the deployment of an up convolution operation, complete with a 2x2 kernel size. This intricate operation ingeniously halves the number of feature channels, striking a harmonious balance between data volume and computational efficiency. An integrative touch arises with the concatenation of the result from the expansive path with the corresponding feature map from the contracting path. This strategic interweaving facilitates the holistic utilization of both high-level and low-level features. This fusion of insights subsequently undergoes a twofold convolution, where each convolution layer sports a 3x3 kernel size, complemented by the ReLU activation function.

The U-Net architecture stands as a compelling testament to innovation within neural network design. Its ingenious fusion of encoder and decoder blocks, coupled with strategic feature manipulation, positions it as a stalwart solution in tasks such as image segmentation and medical image analysis. The U-Net's prowess lies in its ability to seamlessly weave together the finer aspects of feature extraction and pixel classification, culminating in a transformative neural framework.

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Issue: 09 | Sep 2023 www.irjet.net p-ISSN: 2395-0072

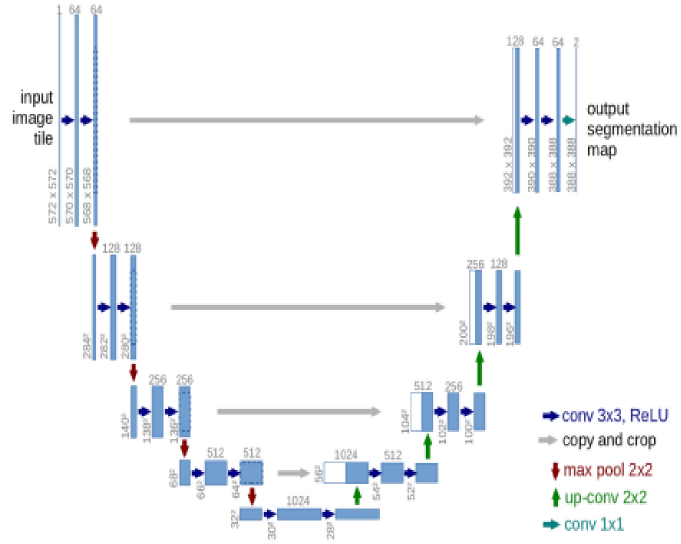


Fig-3: U-Net Architecture Diagram

3.3 ResUnet Architecture

ResUNet is a fully convolutional neural network strategically engineered for exceptional performance while keeping parameter count to a minimum, thus representing an enhancement over the existing UNet architecture. It adeptly leverages the strengths of both the UNet framework and Deep Residual Learning.

The ResUNet comprises three integral components: an encoding network, a decoding network, and a connecting bridge, akin to the U-Net structure. In contrast to the U-Net, where two 3 x 3 convolution layers are followed by ReLU activation functions, ResUNet replaces these layers with pre-activated residual blocks.

The integration of residual blocks serves multiple purposes, including enabling the creation of a deeper network without grappling with issues like vanishing or exploding gradients. Additionally, it simplifies the network's training process.

ResUNet is a modified version of the U-Net architecture, which is a convolutional neural network (CNN) architecture commonly used for image segmentation tasks, such as medical image segmentation. ResUNet incorporates residual connections, inspired by the ResNet architecture, to improve the training and performance of U-Net.

Moreover, within ResUNet, the inclusion of rich skip connections facilitates more efficient information flow among different layers, greatly enhancing gradient flow during the training process (backpropagation).

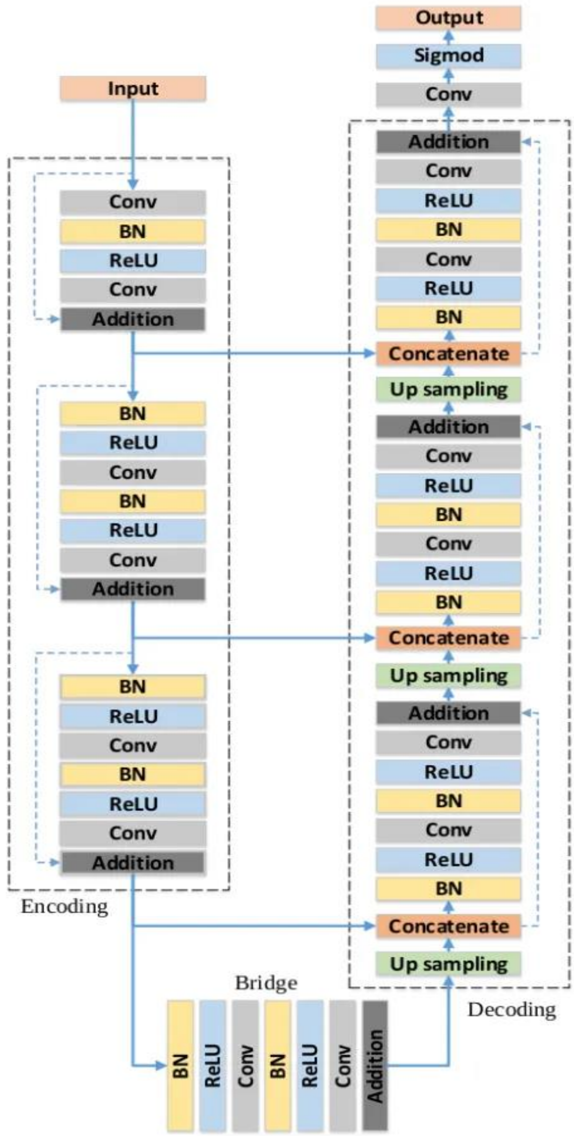


Fig-4: ResUnet Architecture Diagram

3.4 Attention U-Net Architecture

Attention plays a vital role in computer vision tasks, particularly when directing focus to specific image regions becomes necessary. In the context of image segmentation, attention proves immensely valuable as it highlights crucial areas within the image. Due to this the model converges much faster and has better generalization capabilities.

Attention enhances the model's ability to focus on relevant features and improve its segmentation performance. Attention mechanisms have been incorporated into UNet-like architectures to help address issues such as capturing fine details, handling class imbalance, and improving overall segmentation accuracy.

The "Attention UNET" framework introduces an innovative spatial attention mechanism known as the Attention Gate (AG). This Attention Gate plays a pivotal role in focusing on target structures of varying shapes and sizes within medical images.

When trained with the Attention Gate, models inherently learn to suppress irrelevant image regions while accentuating salient features essential for the task at hand. This intrinsic ability eliminates the need for explicit localization modules.

Furthermore, integrating the Attention Gate into existing segmentation architectures, such as U-Net, incurs only a minimal increase in parameters. This parameter boost, in turn, elevates the model's sensitivity and accuracy.

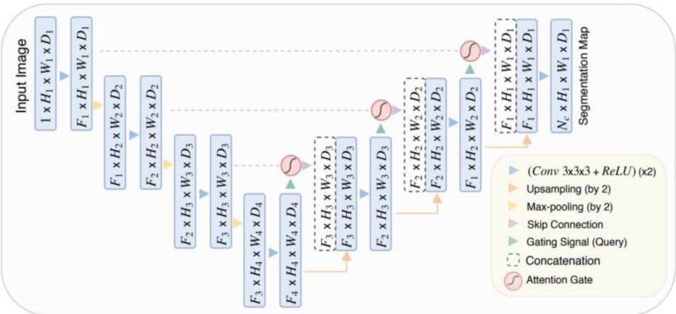


Fig-5: Attention Unet Architecture Diagram

3.5 Datasets

The following class imbalanced datasets have been used for this project.

CVC- Clinic DB- Contains 612 images of Colorectal Polyp. The images are RGB with dimensions 288 x 384 x 3. These images are generated from 23 video sequences obtained from 13 different patients using standard colonoscopy interventions with white light.

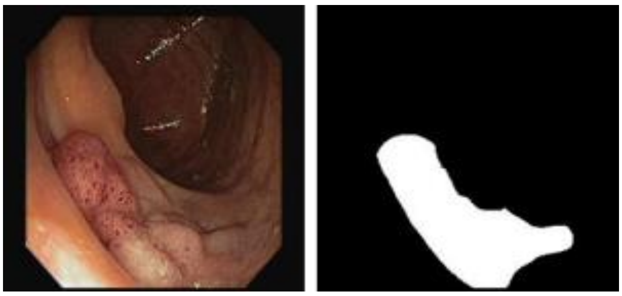


Fig-6: Example of colorectal polyp and its mask

ISIC 2018 Skin Lesion Dataset -The ISIC 2018 Skin Lesion Dataset is a comprehensive collection comprising a total of 2596 images, all of which depict various instances of skin lesions. These images are captured in the RGB color space, providing rich visual information for analysis. Notably, the dataset exhibits diversity not only in terms of the lesions themselves but also in the acquisition process. The images are obtained using different types of dermatoscopes, reflecting variations in imaging techniques and equipment. Furthermore, the dataset draws from multiple medical institutions, ensuring a broad representation of skin conditions and lesions across different patient populations. These images serve as a valuable resource for researchers and medical professionals in the field of dermatology and skin cancer screening, enabling the development and evaluation of advanced image analysis algorithms and diagnostic tools.

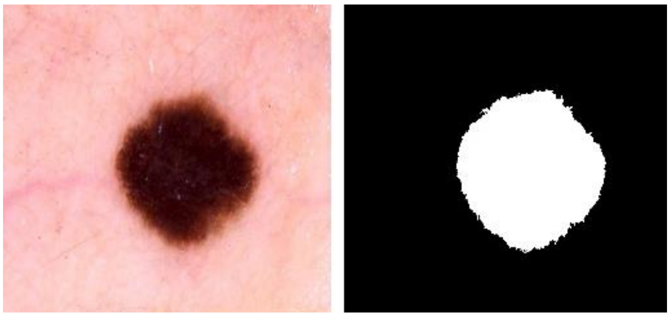


Fig-7: Example of Skin lesion and its mask

Division of datasets into training, validation and testing dataset

Out of the entire dataset, 10 percent of the data is used as validation set and 10 percent data is used as test data.

Table 1- Division of datasets into train, test and validation set

Dataset	Training Images	Validation Images	Test Images
CVC Clinic DB	490	61	61
ISIC 2018 Skin Lesion dataset	2076	259	259

Loss Functions and Metrics used for Evaluation

Four different loss functions namely - Binary cross entropy loss, Focal Loss, Dice Loss and combo Loss have been used for this research. Precision, Recall, Dice score, intersection over union(IoU) are the metrics on which the performance of the model is evaluated.

4. PROPOSED METHODOLOGY

Our proposed method consists of multiple models based on multiple architectures against different loss functions for image segmentation. The methodology encompasses dataset preprocessing, model training, testing, performance evaluation and hyperparameter tuning.

Data Preprocessing: The collected dataset undergoes preprocessing to ensure consistent quality and remove any irrelevant or duplicate images. Common preprocessing steps include resizing images to a uniform resolution, normalizing pixel values, and removing any noise or artifacts.

Training: The preprocessed dataset is then used to train the UNet, ResUnet and Attention Unet models. The model tries to learn the hidden features in the images. It is important that the model learns sufficiently on the dataset since it is class imbalance but not overfit on the dataset. The models are trained for 100 epochs along with early stopping criteria.

Testing: Once the UNet, ResUnet and Attention Unet models are trained on the medical image dataset they are used to generate segmented masks on previously unseen data.

Hyperparameter tuning: The models were trained for 100 epochs along with stopping criteria. The learning rates used were 0.1, 0.01, 0.001. The learning rate was reduced in case of a plateau. The model had an early stopping criteria by monitoring validation loss with a patience of 20 epochs.Throughout this methodology, it's essential to follow best practices in machine learning, such as splitting the dataset into training, validation, and test sets, tuning hyperparameters, and optimizing the model's architecture. Regular monitoring and fine-tuning of the model based on results can lead to improved performance over time.

5. RESULT AND DISCUSSION

5.1 Dataset

The dataset consists of Skin lesions and polyp images collected from various medical institutes. Models like U-net, Attention u-Net, ResUnet are capable of segmenting the image with a good accuracy. Performance is evaluated using metrics such as precision, recall, dice score and IoU(Intersection over union). The model is capable of segmentation to label the regions as tumour and surrounding tissue.

5.2 Convolutional Neural Network

Convolutional Neural Networks (CNNs) have emerged as a fundamental component in the field of medical image segmentation, including the analysis of skin lesion images within the ISIC 2018 Skin Lesion Dataset. CNNs are a class of deep learning models designed to excel in tasks involving grid-like data such as images. They have revolutionized image processing, feature extraction, and pattern recognition due to their inherent ability to automatically learn hierarchical features from raw pixel data.

In the context of medical image segmentation, CNNs have proven particularly effective because of their capacity to capture intricate details and spatial relationships within images. This capability is crucial for accurately identifying and delineating the boundaries of skin lesions, which vary in shape, size, and texture.

CNN architectures consist of multiple layers, including convolutional layers, pooling layers, and fully connected layers. The convolutional layers, in particular, play a pivotal role by applying convolutional operations to the input data. These operations involve sliding small filters (kernels) over the input image to detect local patterns, edges, and features. As the network progresses through successive convolutional layers, it learns to extract increasingly abstract and high-level features.

Furthermore, pooling layers serve to reduce the spatial dimensions of feature maps, enabling the network to focus on the most salient information while maintaining computational efficiency. This hierarchical feature extraction, combined with non-linear activation functions like Rectified Linear Units (ReLU), empowers CNNs to uncover complex patterns within medical images.

For skin lesion segmentation, CNNs are typically employed in a fully convolutional manner, meaning they can accept input images of varying dimensions. This adaptability aligns well with the ISIC 2018 Skin Lesion Dataset, which consists of RGB images with different dimensions. Moreover, CNNs can be fine-tuned and customized to suit specific segmentation tasks, such as distinguishing between skin lesions and the surrounding healthy tissue.

In our research, CNNs were a cornerstone of our image segmentation pipeline, working in synergy with various loss functions and architectural variations. Their robustness and adaptability make them an indispensable tool in the pursuit of accurate and efficient skin lesion segmentation, a critical step in skin cancer diagnosis and screening.

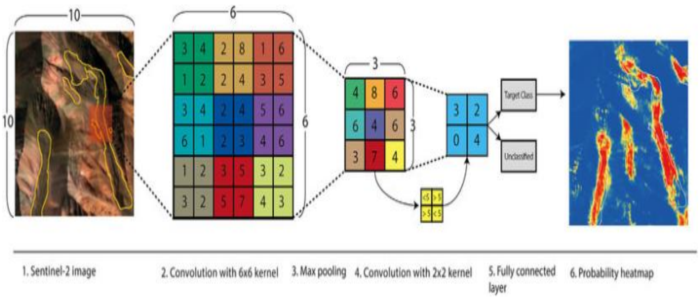


Fig -8: CNN Layers for input image.

5.3 Image Segmentation

Image segmentation involves a series of key steps to transform raw image data into meaningful regions or segments.

1. *Preprocessing:*

At the outset, the image typically undergoes preprocessing to enhance its quality.These preprocessing steps may encompass noise reduction, contrast enhancement, or other image enhancement techniques.

2. *Feature Extraction:*

Next, features relevant to the segmentation task are extracted from the image. These features can encompass a range of characteristics such as color values, pixel intensity, texture patterns, or gradients.

3. *Clustering or Thresholding:*

Simple segmentation techniques, like thresholding, may be employed. Thresholding involves applying a value to the image to separate regions based on pixel values. Pixels with values above the threshold are grouped into one segment, while those below the threshold form another segment. Alternatively, clustering methods, like K-means clustering, can be utilized to group pixels with similar properties together.

4. *Edge Detection:*

For edge-based segmentation, the focus is on identifying boundaries between objects. Edge detectors like the Canny edge detector are utilized to find abrupt changes in intensity, often corresponding to object boundaries.

5. *Region Growing:*

Region growing is a technique that starts with a seed point and gradually expands a region by adding neighboring pixels that meet specific similarity criteria. This process continues until no more similar pixels can be added to the region.

6. *Watershed Segmentation:*

Watershed segmentation treats pixel values in the image as topographical features, where pixels with similar values form valleys and ridges. Watershed lines are then employed to separate regions based on these topographical features.

7. *Graph-Based Segmentation:*

Graph-based approaches represent the image as a graph, with pixels as nodes and edges indicating connections between neighboring pixels. Graph-based segmentation algorithms aim to partition the graph into disjoint subgraphs corresponding to image regions.

8. *Machine Learning-Based Segmentation:*

Deep learning techniques, particularly convolutional neural networks (CNNs), have revolutionized image segmentation.

CNN-based models are trained on extensive datasets to learn features and object boundaries in images. Popular architectures like U-Net, Mask R-CNN, and FCN have been developed for image segmentation tasks.

9. *Post-processing:*

Following segmentation, post-processing steps are often applied to refine the results. This can involve actions like removing small isolated regions, merging adjacent regions, or smoothing boundaries.

10. *Visualization and Analysis:*

Ultimately, segmented regions are visualized using masks or overlays to facilitate comprehension and analysis.

These segmented regions can be leveraged for various computer vision tasks, including object detection, tracking, or in-depth image analysis.

5.4 Results

Image segmentation is performed on 2 datasets – CVC Clinic DB and ISIC 2018 Skin Lesion dataset. In the CVC Clinic DB dataset, the tumour has to be distinguished from the surrounding tissue. The dataset consists of 612 images. The best results were given by Attention U-Net in combination with the combo loss with a dice score of 0.9062 and anIoU score of 0.8325.

The worst performance was that of ResUnet architecture combined with binary cross entropy loss function with a dice score of 0.4933 and an IoU score of 0.3274.

In the ISIC 2018 Skin Lesion dataset the tumour has to be distinguished from the surrounding skin. The dataset consists of 2596 images of skin lesions. The best results were given by Attention u-net architecture in combination with combo loss with a dice score of 0.8879 and anIoU score of 0.8153.

The worst performance was that of ResUnet architecture combined with binary cross entropy loss function with a dice score of 0.7582 and an IoU score of 0.6710.

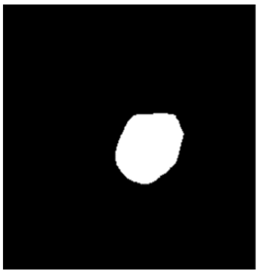
Results on CVC Clinic Dataset (Dataset for Polyp Segmentation)

Fig-9: Image and its ground truth

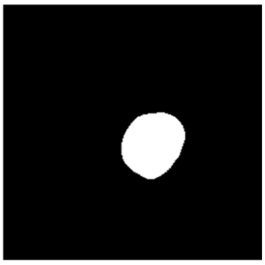
Table 2 - U-Net Architecture

Model	Loss function	Dice Score	IoU	Precision	Recall
U- Net	Binary Cross entropy Loss	0.7531	0.5694	0.9121	0.8491
	Dice Loss	**0.8914**	**0.8041**	0.9022	**0.8687**
	Focal Loss	0.8699	0.7699	**0.9200**	0.8314
	Combo Loss	0.8460	0.7343	0.8793	0.8005

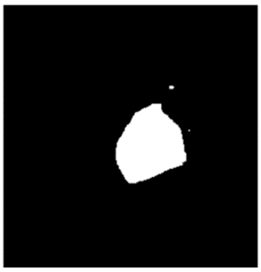
Binary Cross Entropy Loss



Dice Loss



Focal Loss

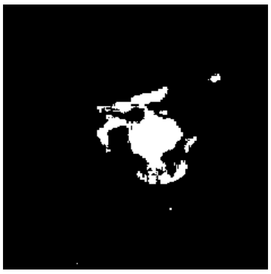


Combo Loss

Fig-10: Segmented Masks

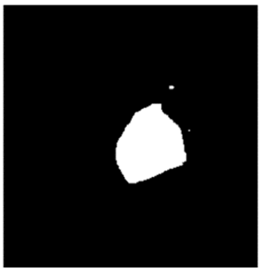
Table 3 - ResU-Net Architecture

Model	Loss function	Dice Score	IoU	Precision	Recall
Res-Unet	Binary Cross entropy Loss	0.4933	0.3274	0.7317	0.4838
	Dice Loss	0.7298	0.5744	0.8445	0.6670
	Focal Loss	0.8124	0.6803	0.8872	0.7428
	Combo Loss	**0.8423**	**0.7279**	**0.9005**	**0.7906**

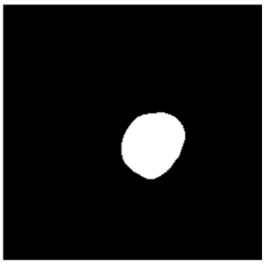


Binary Cross Entropy Loss

Dice Loss



Focal Loss



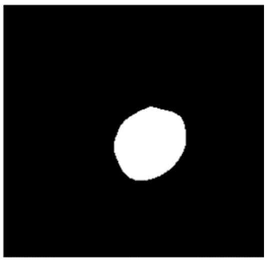
Combo Loss

Fig-11: Segmented Masks

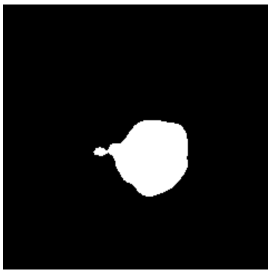
Table 4 - Attention U-Net Architecture

Model	Loss function	Dice Score	IoU	Precision	Recall
Attention U-Net	Binary Cross entropy Loss	0.7821	0.6424	0.9191	0.7946
	Dice Loss	0.8873	0.7214	0.8856	0.8574
	Focal Loss	0.8487	0.7376	0.8919	0.8239
	Combo Loss	**0.9062**	**0.8325**	**0.9488**	**0.8605**

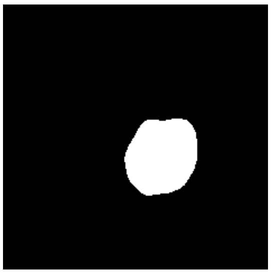
Binary Cross Entropy Loss



Dice Loss



Focal Loss



Combo Loss

Fig-12: Segmented Masks

Results on ISIC 2018 dataset

(Dataset for Skin Lesion Segmentation)

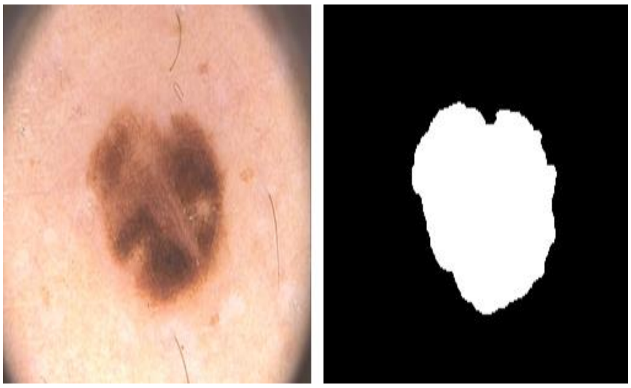
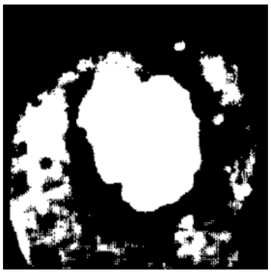


Fig-13: Image and its ground truth

Table 5 - U-Net Architecture

Model	Loss function	Dice Score	IoU	Precision	Recall
U- Net	Binary Cross entropy Loss	0.7689	0.6476	0.7157	0.9092
	Dice Loss	0.8794	**0.8083**	**0.9151**	0.8885
	Focal Loss	0.8778	0.7993	0.8883	0.9021
	Combo Loss	**0.8836**	0.8064	0.8801	**0.9202**



Binary Cross Entropy Loss

Dice Loss

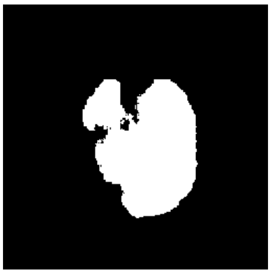
Focal Loss

Combo Loss

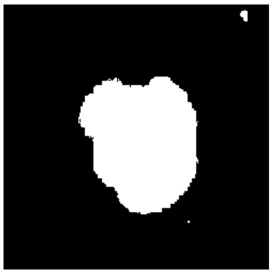
Fig-14: Segmented Masks

Table 6- ResU-Net Architecture

Model	Loss function	Dice Score	IoU	Precision	Recall
Res- Unet	Binary Cross entropy Loss	0.7582	0.6710	0.7109	0.8516
	Dice Loss	0.8273	0.7211	0.7513	0.8874
	Focal Loss	**0.8600**	**0.7866**	**0.8753**	0.9065
	Combo Loss	0.8356	0.7481	0.8365	**0.9083**



Binary Cross Entropy Loss



Dice Loss

Focal Loss

Combo Loss

Fig-15: Segmented Masks

Table 7 - Attention U-Net Architecture

Model	Loss function	Dice Score	IoU	Precision	Recall
Attention U- Net	Binary Cross entropy Loss	0.8398	0.7513	0.8074	0.9375

	Dice Loss	0.8415	0.7482	0.8643	0.8713
	Focal Loss	**0.8879**	**0.8153**	**0.8961**	0.9147
	Combo Loss	0.8770	0.8046	0.8585	**0.9394**

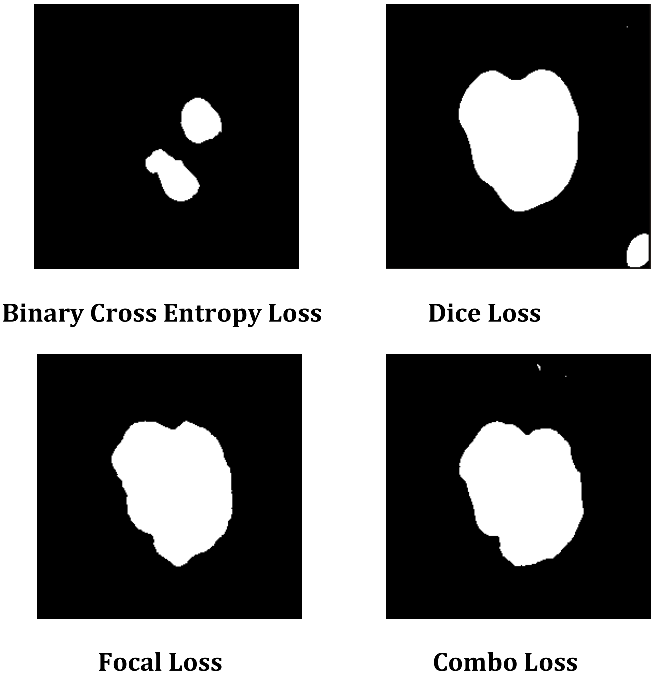


Fig-16: Segmented Masks

6. CONCLUSIONS

This study focused on evaluating the effectiveness of various loss functions using two distinct datasets with class imbalance: CVC Clinic Dataset and ISIC 2018 Skin Lesion dataset. The use of Attention U-Net combined with Combo Loss emerged as a robust solution against class imbalance. Additionally, favorable outcomes were observed in certain cases with Dice Loss and Focal Loss when used with Attention U-Net.

Combo Loss and Focal Loss consistently yielded the most promising results. The disparities in accuracy across different loss functions underscore the significant influence of the chosen loss function on model performance. Conversely, poorer performance was evident when employing distribution-based losses like binary cross-entropy loss.

Thus the combination of Attention U-Net architecture can be used with loss functions such as Combo Loss, Focal Loss to yield the best results since this combination handles class imbalance and produces accurate segmentation results.

Our findings reveal several important insights:

Architecture Matters: The choice of segmentation architecture has a significant impact on the model's performance. Among the architectures studied, Attention U-Net consistently demonstrated superior performance in both datasets. Its incorporation of attention mechanisms allowed it to focus on relevant features, enhancing sensitivity and accuracy.

Loss Functions are Crucial: Addressing class imbalance is essential in medical image segmentation, where tumors often occupy a small portion of the image. Our research showed that traditional loss functions like binary cross-entropy may not be sufficient. Loss functions like Dice Loss, Focal Loss, and Combo Loss, which are more robust to class imbalance, consistently outperformed cross-entropy loss in our experiments.

Dataset-Specific Variations: Different datasets present unique challenges, and the choice of architecture and loss function should be tailored to the specific dataset. For instance, in the CVC Clinic DB dataset, the Attention U-Net with Combo Loss performed exceptionally well, while in the ISIC 2018 Skin Lesion Dataset, the same combination achieved remarkable results.

Clinical Relevance: Accurate medical image segmentation has the potential to make a significant impact on the healthcare industry. Our research contributes to the development of intelligent systems that can assist medical professionals in diagnosing conditions such as colorectal polyps and skin lesions with high precision.

In conclusion, our study underscores the importance of carefully selecting the combination of architecture and loss function for medical image segmentation tasks. Attention U-Net, in conjunction with robust loss functions like Combo Loss, has demonstrated promising results in tackling class imbalance and accurately segmenting tumors. However, further research and validation on larger and more diverse medical imaging datasets are essential to enhance the generalizability and clinical applicability of these findings.

This research serves as a stepping stone towards the development of intelligent systems capable of transforming the medical field by aiding in the early and accurate diagnosis of various medical conditions, ultimately improving patient outcomes and healthcare efficiency.

REFERENCES

[1]Abraham, N., Khan, N.M., 2019. A novel focal Tversky loss function with improved attention U-Net for lesion segmentation. Proc. 16th International Symposium on Biomedical Imaging (ISBI).IEEE,pp.683–687.https://doi.org/10.1109/ISBI.2019.8759329.

[2]Unified Focal Loss: Generalising Dice and cross entropy based losses to handle class imbalance medical image segmentation by M.Yeung, E. Sala, C. Schonlieb , L. Rund

[3] Noel Codella, Veronica Rotemberg, Philipp Tschandl, M. Emre Celebi, Stephen Dusza, David Gutman, Brian Helba, Aadi Kalloo, Konstantinos Liopyris, Michael Marchetti, Harald Kittler, Allan Halpern: "Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC)", 2018; https://arxiv.org/abs/1902.03368

[4] Tschandl, P., Rosendahl, C. & Kittler, H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci. Data 5, 180161 doi:10.1038/sdata.2018.161 (2018).

BIOGRAPHIES

Chevi Parsai is a B.Tech student at SGSITS, Indore in Computer Science and Engineering in the year 2020-2024. Her research area is Machine Learning, Deep Learning, Artificial Intelligence and Computer Vision.

Palak Kakani is a B.Tech student at SGSITS, Indore in Information Technology and Engineering in the year 2020-2024. Her research area is Machine Learning, Deep Learning, Data Science, Data Analysis and Data Visualization.

Shreya Vyas is a B.E student at IET DAVV, Indore in Computer Science and Engineering in the year 2020-2024. Her research area is Machine Learning, Deep Learning, Data Science, Data Analysis and Data Visualization.