Skin Cancer Image Classification Based on Neural Networks

Saeid Moradi

***Abstract*- Skin cancer is a common type of cancer that can be difficult to diagnose accurately, especially in its early stages. Traditional diagnosis methods rely on visual inspection by dermatologists, which can be subjective and limited by the expertise and experience of the clinician. In recent years, artificial neural networks (ANN), including multilayer perceptron (MLP) and convolutional neural networks (CNNs), have shown promise in the automated detection of skin cancer. This study has employed both MLP and CNNs with an image augmentation technique to classify skin cancer images in the HAM10000 dataset. The results of this study show that the MLP and CNNs have achieved accuracies of 72.6 and 78.8, respectively, in classifying skin cancer images. These results demonstrate the potential of ANNs in the automated diagnosis of skin cancer.**

1. Introduction

Skin cancer is among the most prevalent forms of cancer in the current decade [1]. Skin cancer is generally divided into two main categories, namely melanoma and nonmelanoma skin cancer [2]. Melanoma is a particularly dangerous, infrequent, and lethal type of skin cancer. Melanoma cancer originates from melanocytes, which are a type of skin cell. This cancer occurs when healthy melanocytes start to multiply uncontrollably, resulting in the formation of a cancerous tumor. Melanoma can develop in any part of the body but typically occurs in areas that are frequently exposed to sunlight, such as the face, neck, hands, and lips [3]. The major skin cancer types comprise basal cell carcinoma [4], squamous cell carcinoma [5], Merkel cell cancer [6], dermatofibroma [7], vascular lesion [8], and benign keratosis [9].

As melanoma cancer can spread to other parts of the body and result in a painful and fatal outcome for the patient, timely detection is crucial in effectively treating [10]. Nonmelanoma skin cancers typically have a lower tendency to spread to other parts of the body and are generally easier to treat. Consequently, early detection plays a crucial role in the effective treatment of skin cancer.

The usual method for detecting skin cancer is biopsy, which involves the removal of a tissue sample from a suspicious skin lesion for laboratory analysis to determine whether it is cancerous or not. However, this approach can be painful, time-consuming, and costly [3]. Computer-based technology offers a non-invasive, more affordable, and faster method for diagnosing skin cancer symptoms, whether they are classified as melanoma or nonmelanoma. To detect skin cancer, multiple non-invasive techniques are proposed for examining the symptoms. The general process involves capturing an image, preprocessing it, segmenting the preprocessed image, extracting relevant features, and classifying it. In this project Multilayer perceptron and convolutional neural networks are used for classifying skin cancer images.

2. *Data*

We know that quality data plays pivotal roles in the performance of machine learning models. Therefore, to assess the effectiveness of computer-based systems for skin cancer diagnosis, a diverse and comprehensive collection of dermoscopic images is necessary. However, many skin cancer datasets have been limited in size and diversity, often containing only images of one or two lesion types. To address the issue of limited diversity and small size in dermatoscopic image datasets for training neural networks in automated diagnosis of pigmented skin lesions, the HAM10000 dataset has been created [11]. The dataset consists of 10,015 dermatoscopic images obtained from different populations and acquired through various modalities. It includes representative cases of all significant diagnostic categories for pigmented lesions such as actinic keratoses and intraepithelial carcinoma (akiec), basal cell carcinoma (bcc), benign keratosis-like lesions (bkl), dermatofibroma (df), melanoma (mel), melanocytic nevi (nv), and vascular lesions (vasc) [11].

It took twenty years to compile and consists of images sourced from Cliff Rosendahl's skin cancer practice in Queensland, Australia, and the Dermatology Department of the Medical University of Vienna, Austria. The photographic prints of lesions were digitized using a Nikon-Coolscan-5000-ED scanner, and the resulting images were cropped and saved in JPEG format with 300 DPI quality. A semi-automatic workflow was developed to clean the images and ensure diversity using a neural network. The dataset comprises 327 images of AK, 514 images of basal cell carcinomas, 1099 images of benign keratoses, 115 images of dermatofibromas, 1113 images of melanocytic nevi, 6705 images of melanomas, and 142 images of vascular skin lesions [3]. Figure 1 indicates some images for seven different lesion types from the dataset.

More than 50% of the lesions in the dataset are confirmed through histopathology (histo), while the rest have either follow-up examination (follow_up), expert consensus (consensus), or confirmation by in-vivo confocal microscopy (confocal) as the ground truth. The dataset also includes lesions with multiple images that can be tracked using the lesion_id column in the metadata file [11].

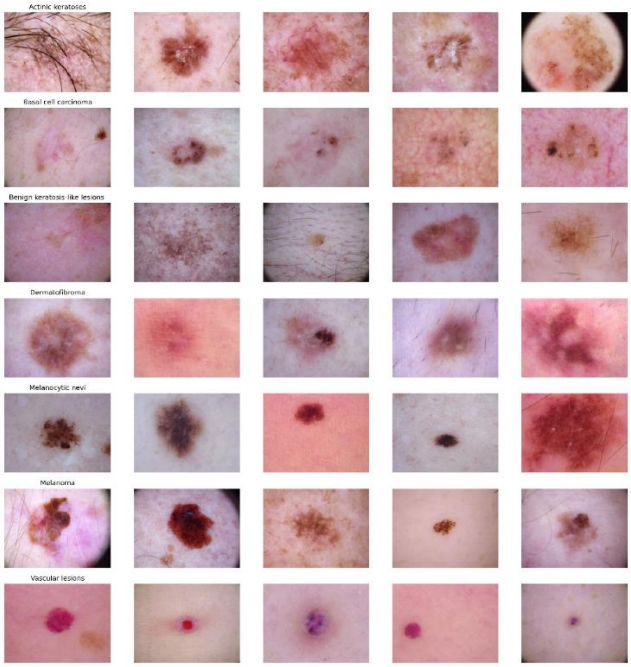

Figure 1. Skin Disease Categories from HAM10000 Dataset [11].

2.1 Exploratory Data Analysis – EDA

Exploratory data analysis as the process of analyzing and summarizing dataset helps to gain a better understanding of dataset characteristics, before formal modeling is performed. The main goal of EDA is to identify patterns, trends, and relationships in the data that can inform further analysis and modeling.

Figure 2 represents some information about the HAM10000 dataset. The bar chart for age groups shows that people in the age range of 45 years old are more likely to suffer from skin diseases as compared to other age groups. Moreover, it has been observed that youngers tend to have a smaller number of skin diseases. Generally, it can be concluded that as people get older, their chances of getting skin diseases tend to increase.

The bar chart for gender indicates that skin diseases are more prevalent among men as compared to women. The body location bar chart shows skin diseases tend to occur more frequently on the back of the body, whereas the acral surfaces (such as the limbs, fingers, or ears) tend to have the lowest incidence of skin diseases. This pattern could be due to differences in skin thickness, exposure to environmental factors, and other physiological factors.

Finally, lesion type bar chart indicates among the various types of skin diseases, melanocytic nevi is the most diagnosed condition among people. On the other hand, dermatofibroma is a benign skin lesion that is less common than other lesion types in dataset.

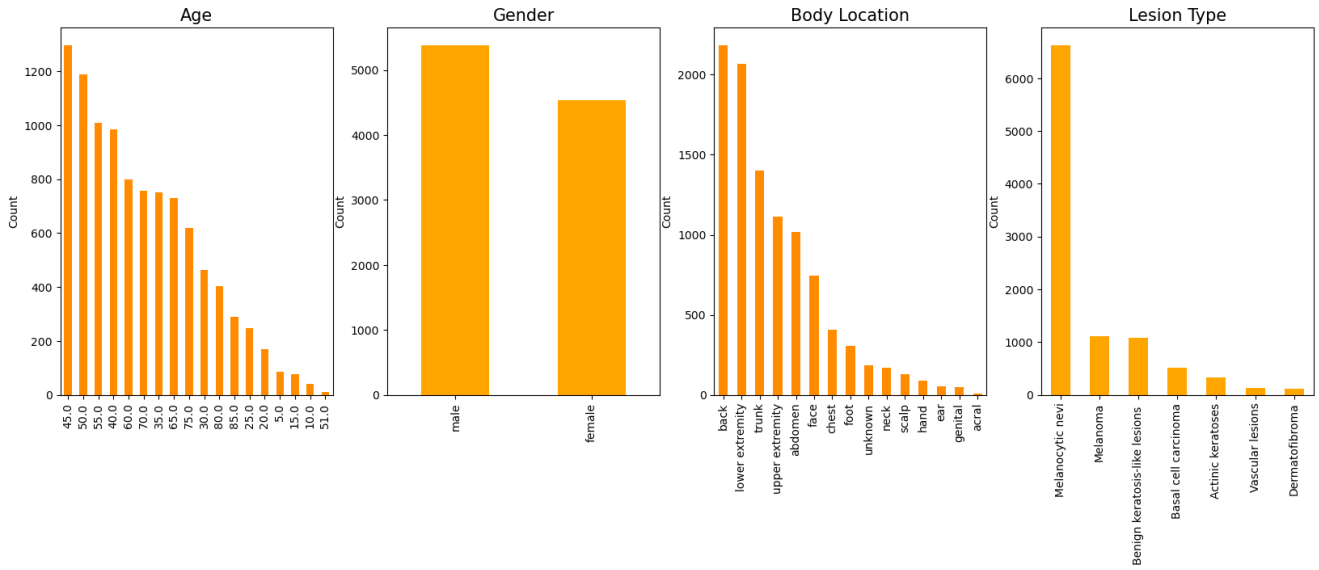

Figure 2. HAM10000 Dataset EDA.

Figure 3 shows skin diseases based on gender. According to the chart, back is the most affected part of the body by skin diseases among men in HAM10000 dataset. On the other hand, the lower extremities, such as the legs, are the most affected part of the body by skin diseases among women. The acral surfaces, which include the fingers, toes, and ears, are the least commonly affected parts of the body by skin diseases, particularly in men.

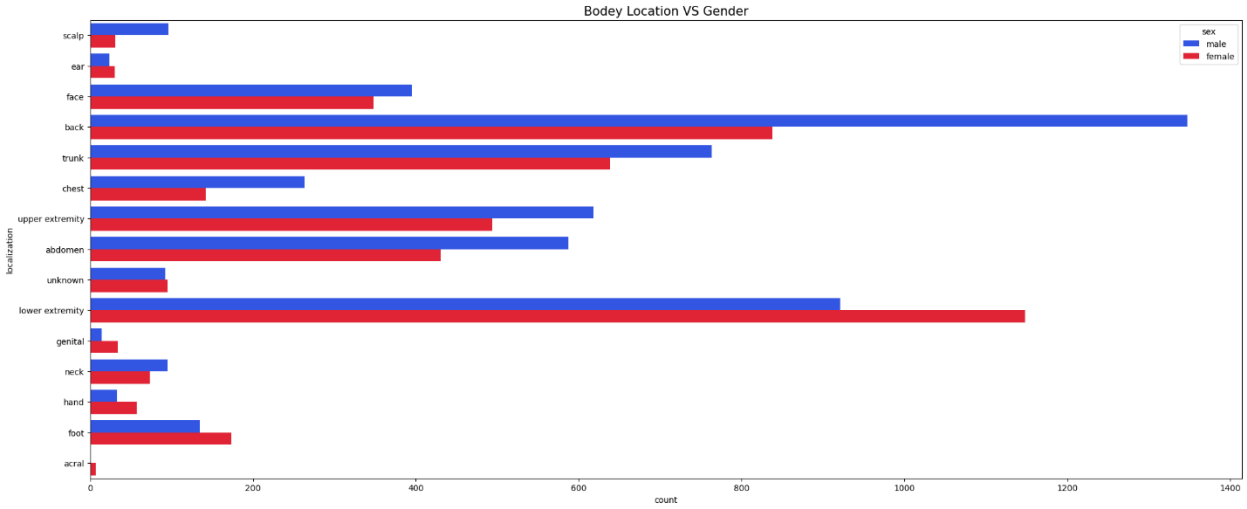

Figure 3. Lesion Types based on Gender in HAM10000 Dataset.

3. *Method*

This study aims to develop an effective skin cancer image classification system using MLP and CNNs. To achieve this goal, we collected a dataset from HAM10000 and preprocessed it before training the models. The original images in the dataset were of size 450*600*3, which is computationally expensive and can cause memory issues during training. Therefore, we resized the images to a smaller size of 150*200*3, which reduces computational load and improves

training efficiency. Additionally, we normalized the images to ensure that the models are fed with standardized data. The preprocessed data was then split into training, validation, and testing sets. With this preprocessed data, we developed and trained neural network models for skin cancer detection.

3.1 Multilayer Perceptron

Multilayer perceptron is a powerful machine learning method that employs nonlinear statistical techniques. This method is inspired by the structure and function of the human brain. The MLP architecture comprises three layers of neurons, where the first layer is called the input layer, which receives data and transfers it to the intermediate or hidden layer of neurons. In a typical MLP, there may be several hidden layers where the intermediate neurons perform computations and transmit the processed data to the output layer, which is composed of output neurons. For training MLPs, backpropagation is the most used method which allows the system to learn complex relationships between the input and output layers. The basic structure of an MLP network is presented in Figure 4.

Choosing the number of hidden layers, number of nodes in hidden layers, learning rate and bath size in an MLP network is a challenge. These are hyperparameters of the models and can be found with cross validation.

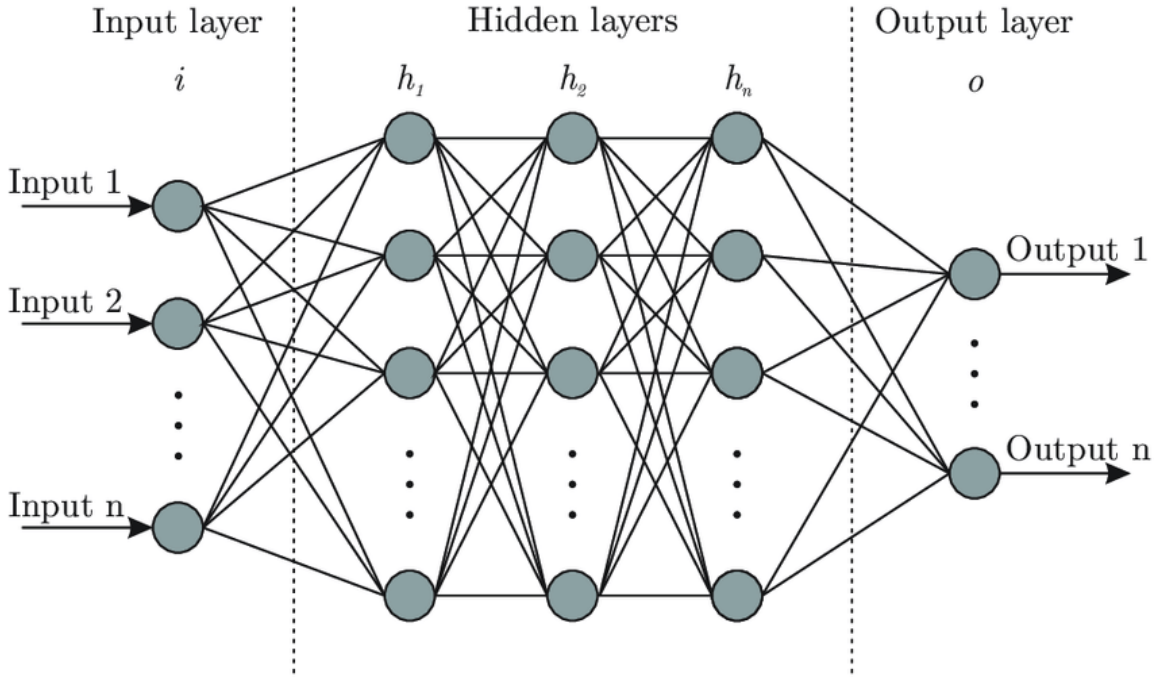

Figure 4. Basic MLP Structure.

3.2 Convolutional Neural Networks

Convolutional neural network is a type of deep neural network which is common in image classification applications. The power of CNN lies in its ability to collect and learn both global

and local data by gathering simpler features such as curves and edges and combining them to form complex features such as shapes and corners [12]. The hidden layers of a CNN typically include convolution layers, nonlinear pooling layers, and fully connected layers [13]. Figure 5 shows the basic architecture of a CNN.

The ability of CNNs in handling large and complex datasets makes them the ideal choice for image classification and recognition tasks. The convolution layer applies filters to input images, which helps in extracting important features such as edges and special patterns from the images. The pooling layer helps in reducing the dimensions of the output of the convolution layer while retaining important features. The fully connected layer processes the output of the convolution and pooling layers, leading to the final output of the network.

Optimizing hyperparameters for CNNs can be a challenging task, similar to MLPs. Key hyperparameters for CNNs include batch size, number of filters, filter sizes, learning rate, and dropout rate. In order to obtain the best model performance, a range of hyperparameter values need to be explored, however, this process can be computationally expensive, making it a challenging issue.

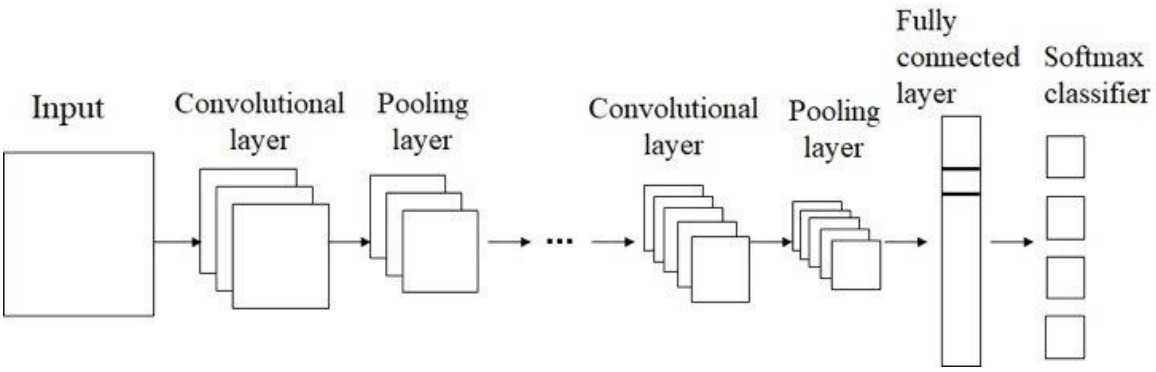

Figure 5. Basic Architecture of a CNN Model.

3.3 Data Augmentation

The HAM10000 dataset exhibits an imbalanced distribution, with some categories having a disproportionately large number of images, while others have only a few. Such imbalances can lead to suboptimal generalization and model overfitting. To mitigate these problems, resampling techniques can be applied. However, in image datasets, where the arrangement and interrelationships of pixels are critical, conventional resampling methods, such as bootstrapping, may not be appropriate. Instead, data augmentation can be utilized to address the issue of imbalanced data. This method involves generating new images by modifying the existing ones using operations such as rotation, flipping, and scaling. This technique can promote dataset balance and enhance model performance.

Data augmentation is a strategy that is commonly utilized in machine learning and computer vision to artificially enlarge the size of a training dataset by generating new, modified versions of the original images. This technique can be applied in various ways, such as rotating, scaling, flipping, cropping, and adding noise to the images. By employing data augmentation, a single image can be altered in multiple ways, such as being randomly rotated, flipped horizontally or vertically, or

having random noise incorporated into it. These modifications lead to numerous versions of the same image with minor alterations. The technique helps to enhance the model's ability to adapt to variations in the input data, such as differences in lighting, angle, or scale, thereby making it more robust.

4. Results

The effectiveness of the suggested MLP and CNN systems was evaluated through experiments conducted on the HAM10000 dataset.

4.1 Parameter Setting and Experimental Evaluation Index

The performance of the suggested neural network systems was evaluated through simulations conducted on the HAM10000 dataset. The simulations were run on a Windows desktop equipped with an AMD Ryzen 7 6800H processor with Radeon Graphics, operating at 3201 Mhz, and having 8 cores and 16 logical processors. The system also had 16GB of RAM. The TensorFlow Keras program was utilized for implementing the proposed scheme. The dataset was split into train, validation, and test sets randomly using a ratio of 80%, 10%, and 10%, respectively.

Table 1 summarizes the values of the hyperparameters utilized by the Adam optimizer for training the proposed architecture. The optimizer implements a learning rate strategy that decelerates learning when it remains stagnant for a prolonged time, known as validation patience. The hyperparameters include the learning rate, the number of epochs, the batch size, and the validation patience. These values are critical for obtaining accurate results while minimizing the computational expense of training the model.

Table 1. Adam Optimizer Parameters

Parameters	MLP Values	CNN Values
Learning rate	0.0008	0.001
Loss function	Categorical Cross entropy	Categorical Cross entropy
Momentum	0.9	0.9
Batch Size	32	16

4.2 Performance Assessment

In this section of the research, we provide a detailed explanation of the evaluation metrics used in the study and their corresponding results. The primary evaluation metric utilized for measuring the effectiveness of the classification model was the classifier accuracy (Acc), which is calculated as the number of accurately classified images divided by the total number of images in the dataset. Table 2 indicates the accuracy for MLP and CNN models. It is obvious that CNN model by 0.7884 accuracy on test data has better results than MLP which has accuracy of 0.72608 for skin cancer image classification.

Table 2. Accuracy MLP and CNN Methods.

Method	Validation Data Accuracy	Test Data Accuracy
MLP	0.7261	0.72608
CNN	**0.7727**	**0.7884**

5. Conclusion

The study proposed the use of MLP and CNNs for the diagnosis of different types of skin diseases by analyzing images of skin lesions from the HAM1000 dataset. The proposed system utilized an image augmentation approach to prevent overfitting, and both models were trained on 80% of the dataset as training data and their performances were evaluated on the remaining 10% validation and 10% test data. The results showed that the CNNs model achieved a higher overall accuracy rate of 78.84% compared to the MLP model's accuracy of 72.61%.

During the project we encountered the issue of computational complexity in neural networks for image dataset, and one possible solution can be using cloud-based systems. Another suggestion is the use of transfer learning methods that involve pre-trained models such as DenseNet, VGG, or AlexNet to analyze more complex datasets with a larger number of cancer cases.

To further evaluate the effectiveness of the proposed method, the study suggested conducting tests on a more diverse dataset that includes images of people with different skin tones from various regions around the world. This would help to ensure that the system is robust and accurate in diagnosing skin diseases in people with different skin tones and from different geographical regions. Overall, the study's results and suggestions provide valuable insights into the use of neural networks for skin disease diagnosis and highlight the need for further research in this area.

The GitHub link of the project is available on the following link:

https://github.com/saeid436/DASC5420-FinalProject

References

[1] Ashraf, R.; Afzal, S.; Rehman, A.U.; Gul, S.; Baber, J.; Bakhtyar, M.; Mehmood, I.; Song, O.Y.; Maqsood, M. Region-of-Interest Based Transfer Learning Assisted Framework for Skin Cancer Detection. IEEE Access 2020, 8, 147858–147871.

[2] Elgamal, M. Automatic Skin Cancer Images Classification. IJACSA 2013, 4.

[3] Dildar, M., Akram, S., Irfan, M., Ullah Khan, H., Ramzan, M., Rehman Mahmood, A., Ayed Alsaiari, S., Hakeem M Saeed, A., Olaythah Alraddadi, M., & Hussen Mahnashi, M. (2020). Skin Cancer Detection: A Review Using Deep Learning Techniques. International Journal of Environmental Research and Public Health. https://doi.org/10.3390/ijerph18105479.

[4] Fuzzell, L.N.; Perkins, R.B.; Christy, S.M.; Lake, P.W.; Vadaparampil, S.T. Cervical cancer screening in the United States: Challenges and potential solutions for underscreened groups. Prev. Med. 2021, 144, 106400.

[5] Ting, D.S.; Liu, Y.; Burlina, P.; Xu, X.; Bressler, N.M.;Wong, T.Y. AI for medical imaging goes deep. Nat. Med. 2018, 24, 539–540.

[6] Wolf, M.; de Boer, A.; Sharma, K.; Boor, P.; Leiner, T.; Sunder-Plassmann, G.; Moser, E.; Caroli, A.; Jerome, N.P. Magnetic resonance imaging T1-and T2-mapping to assess renal structure and function: A systematic review and statement paper. Nephrol. Dial. Transplant. 2018, 33 (Suppl. S2), ii41–ii50.

[7] Hooker, J.M.; Carson, R.E. Human positron emission tomography neuroimaging. Annu. Rev. Biomed. Eng. 2019, 21, 551–581.

[8] Jaiswal, A.K.; Tiwari, P.; Kumar, S.; Gupta, D.; Khanna, A.; Rodrigues, J.J. Identifying pneumonia in chest X-rays: A deep learning approach. Measurement 2019, 145, 511–518.

[9] Morawitz, J.; Dietzel, F.; Ullrich, T.; Hoffmann, O.; Mohrmann, S.; Breuckmann, K.; Herrmann, K.; Buchbender, C.; Antoch, G.; Umultu, L.; et al. Comparison of nodal staging between CT, MRI, and [18F]-FDG PET/MRI in patients with newly diagnosed breast cancer. Eur. J. Nucl. Med. Mol. Imaging 2022, 49, 992–1001.

[10] Khan, M.Q.; Hussain, A.; Rehman, S.U.; Khan, U.; Maqsood, M.; Mehmood, K.; Khan, M.A. Classification of Melanoma and Nevus in Digital Images for Diagnosis of Skin Cancer. IEEE Access 2019, 7, 90132–90144.

[11] Tschandl, P.; Rosendahl, C.; Kittler, H. The HAM10000 Dataset, a Large Collection of Multi-Source Dermatoscopic Images of Common Pigmented Skin Lesions. Sci. Data. 2018, 5, 180161.

[12] Rehman, M.; Khan, S.H.; Danish Rizvi, S.M.; Abbas, Z.; Zafar, A. Classification of Skin Lesion by Interference of Segmentation and Convolotion Neural Network. In Proceedings of the 2018 2nd International Conference on Engineering Innovation (ICEI), Bangkok, Thailand, 5–6 July 2018; pp. 81–85.

[13] Harley, A.W. An Interactive Node-Link Visualization of Convolutional Neural Networks. In Advances in Visual Computing; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2015; Volume 9474, pp. 867–877.