JORMA International Journal of Health and Social Sciences
ISSN: 2978-0942 (Online)
Volume 3 No. 4 Issue 4 Pages 1-16

JORMA IJHSS
DOI: https://doi.org/10.64074/dedfpe19

A Deep Learning-Based Skin Cancer Detection with ResNet50V2

Raphael Ibraimoh[1] || Danial Saraee[2] || Kaveh Kiani[3] || Mohammed Saraee[4]

[1,3,4]School of science, Engineering & Environment, University of Salford, Manchester, M4 5WT, Lancashire, United Kingdom.

[2]Cancer Research, Northern Lincolnshire and Goole NHS Foundation Trust, Scunthorpe, DN15 7BH, Scunthorpe, United Kingdom.

Correspondence:
Dr. Raphael Ibraimoh
Email: r.ibraimoh2@salford.ac.uk.

Abstract

Skin cancer is a prevalent global health concern, with melanoma being particularly dangerous due to its potential to metastasize. Researchers have harnessed deep learning techniques, specifically transfer learning, to create an automated classification system for skin lesions. This system can be especially valuable in areas with limited medical resources. The study leverages pre-trained convolutional neural networks (CNNs) and transfer learning. Using Google Collab and a dataset attached via Google Drive; skin lesion images were analysed. The dataset from International Skin Imaging Collaboration (ISIC) was divided into eight classes, with "Melanocytic Nevus" being the most common (51% of data). Augmentation was applied to address overfitting. Comparing the sequential CNN and pre-trained ResNet50V2 models, both achieved over 60% accuracy. While ResNet50V2 had slightly better accuracy, the sequential model exhibited greater stability in validation results. Notably, segmentation techniques were not employed due to image-specific challenges. This research contributes to improving skin cancer diagnosis and underscores the potential of AI in healthcare. Doctors can benefit from this system, enhancing patient care and treatment decisions.

Keywords: ResNet50V2 || Graphics processing unit || Tensor processing unit || Deep Convolutional Neural Network || Convolutional neural network || Rectified linear unit.

JORMA International Journal of Health and Social Sciences
ISSN: 2978-0942 (Online)
Volume 3 No. 4 Issue 4 Pages 1-16

Introduction

Prolonged exposure to UV radiation, which has been constantly increasing over the years, is the leading cause of skin cancer (Park et al., 2020). Skin cancers develop from abnormal skin cells that can invade other parts of the body. There are three types of skin cancer: basal-cell carcinoma (BCC), squamous-cell carcinoma (SCC), and melanoma. Nonmelanoma skin cancer (NMSC) includes basal cell carcinoma (BCC), squamous cell carcinoma (SCC), and a few less prevalent variants. Basal cell carcinoma (BCC) typically grows slowly and has the potential to cause damage to surrounding tissues. However, it does not typically metastasis to distant locations or provide a significant risk of death. It typically presents as a painless raised skin lesion with a shiny surface and small vessels, or as an elevated area with an ulcer.

Squamous cell carcinoma (SCC), on the other hand, is more likely to spread to other parts of the body and typically appears as a solid mass with a rough, scaly surface that can later evolve into an open sore. Melanomas are the most dangerous type of skin cancer, characterized by changes in the size, configuration, pigmentation, irregular boundaries, multiple hues, pruritus, or bleeding of a mole (Park et al., 2020). Exposure to UV radiation from the sun is responsible for over 90% of skin cancer cases, increasing the risk of acquiring the three main types. This risk has increased, in part due to ozone layer loss. Tanning beds are one of the most common sources of UV radiation exposure. UV exposure throughout childhood is extremely damaging to melanomas and basal-cell cancers. Nonetheless, in the case of squamous-cell skin cancer, the total amount of exposure, regardless of timing, is more important. 20-30% of melanomas arise from preexisting moles. Individuals with fairer skin tones are more vulnerable, as are those with compromised immune systems caused by medicines or disorders such as HIV/AIDS. Typically, the diagnosis is confirmed with a biopsy. Skin cancer is the most common type of cancer worldwide, accounting for at least 40% of all occurrences. Most of these cases are classed as nonmelanoma skin cancer, which affects about 2-3 million people each year. Nonetheless, it is critical to recognize that these figures are imprecise due to little statistical surveillance. Nonmelanoma skin cancer consists of approximately 80% basal-cell tumours and 20% squamous-cell skin (Aima & Akhilesh, 2019).

In the United States, they constitute fewer than 0.1% of the total number of deaths caused by cancer. By comparison, melanoma, a very malignant skin cancer, was detected in over 232,000 people globally in 2012, leading to 55,000 deaths. It is worth mentioning that countries with mostly white populations, such as Australia, New Zealand, and South Africa, have the highest incidence rates of melanoma globally. In the past 20 to 40 years, there has been a rise in the occurrence of the three main forms of skin cancer, especially in areas where most of the population is white. Codella et al. (2018) assert that using machine learning can improve the early detection of strokes, and combining machine learning interpretability libraries can improve both the accuracy of these predictions and the clarity of the resultant conclusions.

Background Study

According to Hassani et al. (2020), AI is divided into three categories: artificial limited intelligence, general AI, and arti- facial superintelligence. Narrow artificial intelligence (AI) exhibits extraordinary proficiency in carrying out specific tasks with intelligence, making it the most common type of AI. In contrast, general artificial intelligence can carry out a variety of cognitive tasks that humans carry out. Artificial superintelligence outperforms human cognitive ability across a wide range of tasks.

Introna (1997) proposed a deep neural network structure to classify dermoscopic images. They conducted their investigation using the ISIC-2019 and ISIC-2020 databases. To address these challenges and obtain remarkable classification performance, we used Efficient Net architecture in conjunction with transfer learning techniques. This approach excels at understanding complicated and nuanced patterns in lesion images by adaptively modifying network depth, width, and resolution. To address the issue of class imbalance, we used dataset augmentation techniques and added metadata to increase the accuracy of our classification results. In addition, to boost the performance of the Efficient Net, we used the ranger optimizer, which simplifies the typically complex process of tweaking hyperparameters required to get cutting-edge results. Marriam et al. (2025) used a transfer learning technique with pre-trained deep CNN models, including ResNET50V2 and VGG16. They propose that the fundamental goal of classifying skin for melanoma and moles is to distinguish between the two, as they are similar. The proposed method employs transfer learning via a pre-trained deep learning network known as ResNET50V2. The Kaggle dataset was used for training and testing, resulting in an impressive classification accuracy of 98.63% in distinguishing between benign and malignant cases. This method outperforms other techniques, including VGG16, which has a classification rate of 97.41%. The results highlight ResNET50V2's outstanding performance, demonstrating its usefulness in photo categorization tasks. Transfer learning can increase model accuracy and efficiency by using knowledge from pre-trained networks, especially in medical image processing (Du-Harpur, Watt, Luscombe, Lynch, 2020).

In a single effort, Haenssle et al. (2018) used a sizable dermoscopy dataset with over 100,000 benign and melanoma lesions. InceptionV4, a deep learning algorithm, was trained and compared to the performance of 58 dermatologists. The diagnosis was divided into two levels: the first used simply dermoscopy, while the second used clinical data and patient photos. Dermatologists showed a median sensitivity of 86.6% and specificity of 71.3% at level I. At level II, the sensitivity and specificity increased to 88.9% and 75.7%, respectively. The improvement in specificity was statistically substantial ($p < 0.05$). Although there was no statistically significant improvement in sensitivity ($p = 0.19$), the Convolutional Neural Network (CNN) based on deep learning demonstrated much better specificity than dermatologists in both level I ($p < 0.01$) and level II ($p < 0.01$). Haenssle et al. (2018) found that CNNs outperformed dermatologists in detecting melanoma through dermoscopy pictures. A Vector Machine (SVM) algorithm with image processing techniques to detect skin cancer was suggested by Roshan (2022). The pre-processing involves noise removal and image enhancement. The segmentation threshold method is used, with the Gray Level Co-occurrence Matrix (GLCM) methodology used for feature extraction. The images were taken using a dermatoscope. Pre-processing consists of three steps: greyscale conversion, noise removal, and image enhancement, which are conducted in that order. Next, the segmentation approach is used to extract the desired region of the image. Figure 1 depicts the block diagram for the entire method.

Figure 1. *Different stages for skin cancer detection.*

JORMA International Journal of Health and Social Sciences
ISSN: 2978-0942 (Online)
Volume 3 No. 4 Issue 4 Pages 1-16

Deep Learning Model Selected

The objective of this research is to develop an advanced multiclass classifier model utilizing deep learning methodologies. We will train the model in a step-by-step manner over multiple epochs, adjusting different parameters to get the best configuration for our specific purpose. Throughout each phase, we will evaluate and verify the models, examining their performance to select the most effective one and comparing it with the transfer learning approach. Model optimization is to optimize the performance of our model at each epoch; we will improve it by running experiments with different parameter values. This technique allows us to select the most suitable model with great accuracy, helping us understand how different combinations of parameters affect their behaviour.

ResNet50V2

ResNet, also known as Residual Network, is a specific variant of convolutional neural network (CNN) that was first proposed in the influential academic paper "Deep Residual Learning for Image Recognition" by He et al. (2016). Convolutional Neural Networks (CNNs) play a crucial role in computer vision applications by facilitating the processing and interpretation of visual data by machines. ResNet-50 is a notable architecture within the ResNet family, consisting of a total of 50 layers. These layers include 48 convolutional layers, one MaxPool layer, and one average pool layer. The distinguishing characteristic of Residual Neural Networks is in their distinct methodology for building deep neural networks. To properly train very deep networks, they utilize residual blocks, each of which is provided with skip connections. Figure 2 shows the architecture of ResNetV2.These connections help the network avoid certain layers during training. This fixes the problem of vanishing gradients and makes it possible to train very complex networks without affecting their performance.

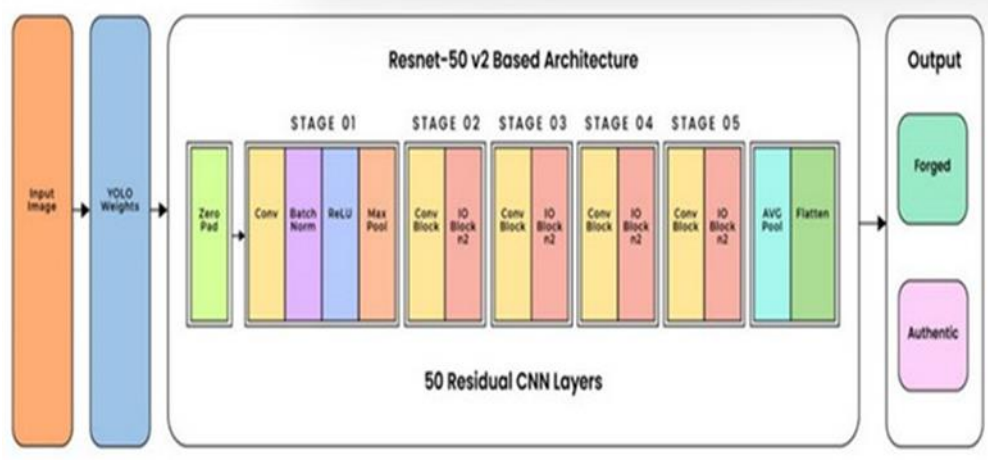

Figure 2. *ResNet50V2 Architecture.*

These connections help the network avoid certain layers during training. This fixes the problem of vanishing gradients and makes it possible to train very complex networks without affecting their performance. ResNet-50 and Residual Neural Networks are significant breakthroughs in the field of Convolutional Neural Networks (CNNs). The authors use novel approaches, such as skipping connections within residual blocks, to address the depth-related constraints that impede

conventional neural net- works. The introduction of these architectural advancements has completely transformed the field of image identification and other computer vision applications. These advancements have allowed for the development of more complex and robust neural networks, resulting in significant enhancements in accuracy and performance (Poonkuzhali, Barathi, & Vinodhkumar, 2022).

Methodology

The primary procedures entail employing a Model Driven Architecture (MDA) tool, such as Domain Specific Languages (DSLs). However, it should be noted that DSLs may not be universally acknowledged as an acronym or abbreviation in the context of neural networks. The construction of a neural network model is facilitated by a user-friendly drag and drop development technique. The abbreviation MDA stands for Model Driven Architecture, a paradigm employed in software development. It will assist in the handling of intricacy, promote communication among stakeholders, and enhance reusability. In summary, the fundamental processes can be outlined as follows.

Data Preparation and Exploratory Data Analysis

The American Cancer Society projects that there will be 97,610 new cases of melanoma in 2023. During this period, a total of 1,958,310 new cases of cancer were recorded, resulting in 609,820 fatalities. According to the World Health Organization (WHO), there are an estimated 2-3 million cases of non-melanoma skin cancers and 132,000 cases of melanoma skin cancers that occur globally each year.

According to the Skin Cancer Foundation, the likelihood of developing cancer is estimated to be 20% for one out of every five Americans. The dataset categorizes skin into eight classes based on its physical features. Every dataset entry contains substantial information about the occurrence of different cancer classes. This study employs a meticulously regulated dataset for scholarly machine learning. The dataset contains 25,333 dermatoscopic pictures that have been meticulously curated to provide comprehensive coverage of pigmented lesions. Our meticulously chosen images cover all significant diagnostic criteria for pigmented lesions, offering a comprehensive and varied resource for machine learning research. The dataset has eight distinct classifications that represent various categories of pigmented lesion diagnoses. The categories were meticulously selected to encompass a broad spectrum of skin problems, thereby furnishing researchers with a comprehensive and representative dataset. This dataset is a valuable asset for academic machine learning, including a diverse collection of dermatoscopic images of pigmented lesions. The designed architecture and inclusion of major diagnostic categories make this tool indispensable for training and testing machine learning algorithms to enhance skin disease identification. The dataset includes the following types of lesions: Actinic Keratosis, Basal cell Carcinoma, Benign Keratosis, Dermatofibroma, Melanocytic Nevus, Melanoma 20, Squamous cell Carcinoma, and Vascular Lesion. Histopathology confirms that over 50% of the remaining patients undergo various examinations, with expert consensus or in-vivo confocal microscopy confirmation being considered the most reliable method.

Image Description

Diagnostic techniques such as clinical examination, dermoscopy, and histological imaging are employed to identify and classify skin conditions. Mobile devices are utilized for capturing

JORMA International Journal of Health and Social Sciences
ISSN: 2978-0942 (Online)
Volume 3 No. 4 Issue 4 Pages 1-16

clinical images to facilitate remote diagnosis and for the purpose of maintaining medical records. They exhibit dermatological issues. Nevertheless, the utilization of dermoscopy and histological images is crucial in determining the severity of skin diseases for clinical diagnosis. This study examines dermoscopy images at the centre like in the figure3. Dermoscopy images, when used with specialized instruments, enable meticulous examination of pigmented skin lesions. Healthcare professionals can utilize high-resolution and magnified photos to precisely analyze skin problems and assist in making correct diagnoses.

Therefore, dermoscopy images are utilized to investigate and evaluate skin disorders in this study. This study used images depicting various fictional individuals with distinct types of skin cancer. The photographs are initially displayed in RGB format, arranged in an 8x6 grid, but the actual images have dimensions of 500x500 pixels. A single representative image is generated for each category to provide a concise depiction of the skin malignancies being examined. This approach streamlines the process of examining datasets and enables focused analysis of different types of cancer and their characteristics.

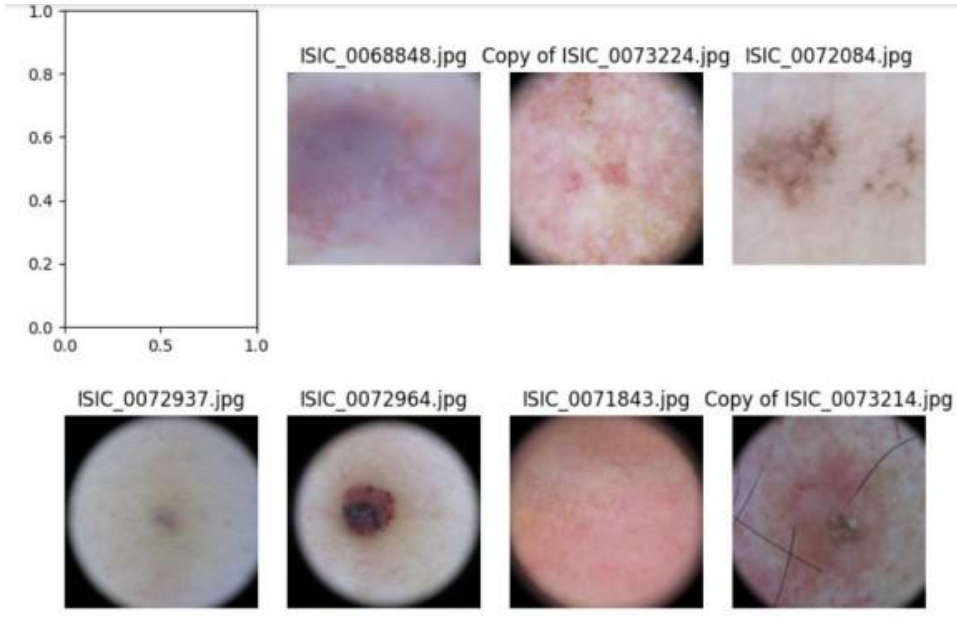

Figure 3. *Image Sample of Class.*

The initial photos in this dataset possessed high resolution and consistent dimensions. Nevertheless, their dimensions were intentionally diminished to align with the objectives of this study. The photographs were resized to dimensions of 100x100 pixels to ensure uniform width and height. This scaling method standardizes images for efficient handling and minimizes the computer resources required for image processing. Due to the computational demands of the neural network employed in this study, the size of the images was diminished. Resizing pictures decreases the computational burden on the network, hence accelerating processing. This deliberate modification guarantees seamless research while maintaining image quality, facilitating effective and successful analysis of the dataset.

Data Wrangling and Transformation

JORMA International Journal of Health and Social Sciences
ISSN: 2978-0942 (Online)
Volume 3 No. 4 Issue 4 Pages 1-16

We employed comprehensive data manipulation and visualization techniques to enhance our comprehension of the dataset's characteristics. An essential stage involved analysing the distribution of skin cancer lesions in the dataset. This study provided valuable information about the distribution of lesion categories in the dataset. Upon thorough analysis, it was found that the "Melanocytic nevus" category had the highest occurrence, with a total of 12,875 cases.

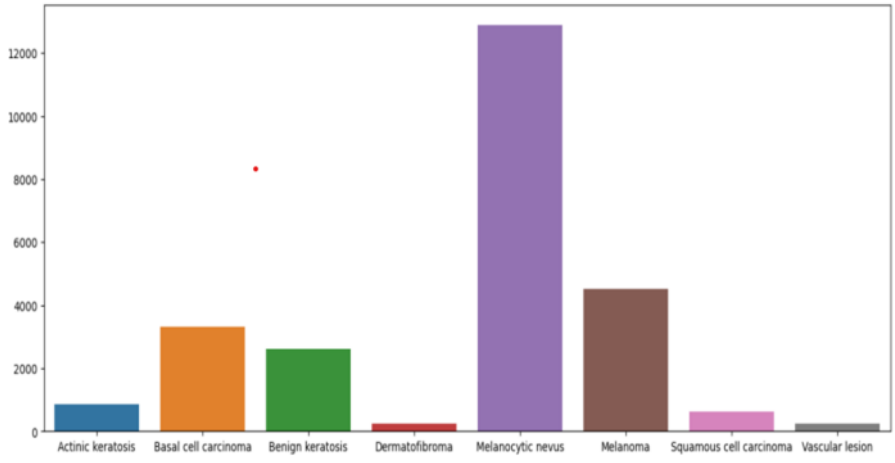

Figure 4. *Bar Chart for total Classes .*

The prevalence of this sort of lesion in the sample indicates its high frequency. Out of all the kinds of lesions, "Dermatofibroma" had the fewest occurrences. The class "Melanoma" had the second-highest occurrence rate, with a total of 4,522 instances. The significance of this group of lesions in the dataset was emphasized as it was the second most common kind of skin cancer. An in-depth examination of skin cancer types revealed that "Melanocytic nevus" had the highest occurrence rate, followed by "Melanoma," while "Dermatofibroma" was the least frequently observed, see figure 4.These findings improved understanding of the dataset and provided valuable insights for future studies.

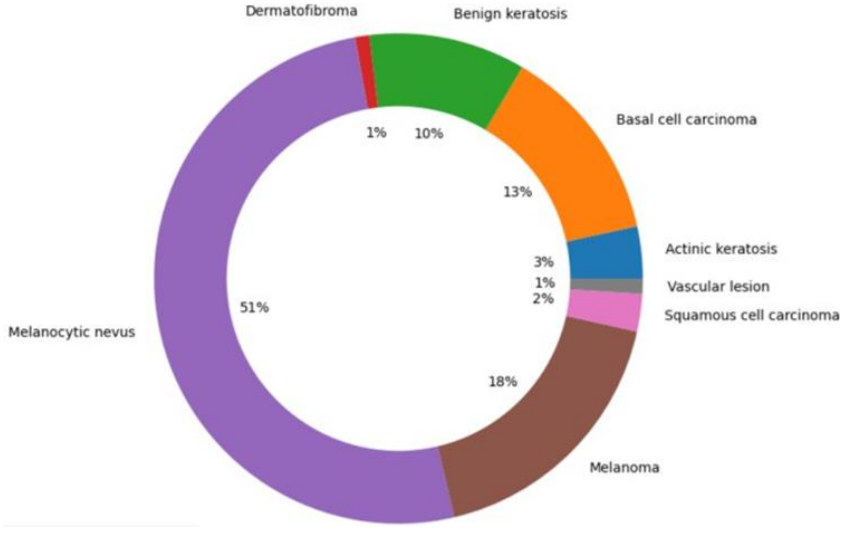

Figure 5. *Donut chart for the Classes Images Distribution .*

The figure 5 shows a considerable data difference in class distribution in this dataset which is highly imbalanced. This observation is especially significant given the "Melanocytic nevus" class dominates 51% of the dataset. In terms of representation, this class dominates the dataset. At 18% of the dataset, the "Melanoma" class is smaller but still significant. At 13% and 10%, respectively, the "Basal cell carcinoma" and "Benign keratosis" classes follow closely. Compared to "Melanocytic nevus," these three classifications are more evenly distributed, although still skewed. Further analysis shows that "Actinic keratosis" makes up only 3% of the sample and "Squamous cell carcinoma" 2%. Last, the "Vascular Lesion" class has the lowest prevalence at 1%. The dataset shows a significant class imbalance, with "Melanocytic nevus" leading, followed by "Melanoma," "Basal cell carcinoma," and "Benign keratosis." The remaining classifications, "Actinic keratosis," "Squamous cell carcinoma," and "Vascular lesion," have smaller shares. Understanding class distribution is vital for researchers and practitioners, emphasizing the need to handle imbalanced data when using machine learning techniques and making data-driven decisions.

Feature Engineering

One of the initial stages in our feature engineering process involves converting picture data into an array format that is compatible with our machine learning model and can be seamlessly incorporated into it. The Python Imaging module (PIL) is used for this crucial transition, although in this case, the NumPy module is employed for the conversion. This library facilitates the conversion of images into arrays, so ensuring their compatibility with the input specifications of our model. The augmentation approach is used to the dataset once the photos have been transformed into arrays. This is done exclusively to achieve a precise outcome and reduce the likelihood of the model overfitting.

Data Augmentations

Data augmentation is a technique that involves creating modified versions of existing data to artificially expand the size of the training set, see figure 6. Another valuable application of this technology is the generation of new data points by updates to the dataset or the utilization of deep learning. A novel term that has emerged in recent years is synthetic data. The contrast between the two of them is as follows. When conducting data augmentation, it is important to make certain minor alterations to the original dataset. These alterations may involve implementing geometric and chromatic adjustments, such as mirroring, resizing, cropping, and fine-tuning the brightness or contrast. The aim of this approach is to enhance the quality of the training dataset by augmenting its diversity while maintaining the fundamental characteristics of the dataset intact (Alavi, 1994). Due to its ability to let machine learning models learn from a wider variety of variations within the same data, it is particularly efficient in cases when the original dataset is either limited in size or lacks diversity.

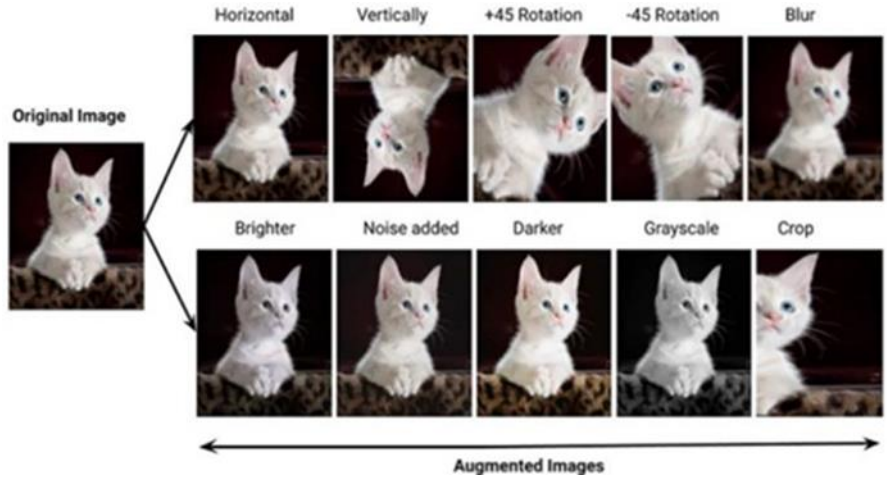

Figure 6. *Augmented Image.*

In contrast, synthetic data production entails generating entirely novel data samples that are not drawn from the original gathered dataset. This process often utilizes advanced machine learning methods, such as Deep Neural Networks (DNNs) and Generative Adversarial Networks (GANs), to create data instances that may not exist but possess the necessary properties or patterns. Models can be trained with data that exceeds the limitations of the original dataset, thus enhancing their performance and resilience. Producing synthetic data offers versatility and can assist in tackling issues associated with limited data availability. Augmentation is a highly successful technique for mitigating the danger of model overfitting and addressing the issue of insufficient initial training datasets. Furthermore, it enhances the accuracy of the model and decreases expenses associated with the annotation and refinement of raw information. The limitations of augmented images include: the supplemented data inherits the same inherent biases as the original dataset, ensuring the quality of augmented data requires expensive quality assurance methods, developing advanced systems like GANs for high-resolution imaging demands extensive research and development efforts, and devising an effective data augmentation plan may pose challenges.

Image Segmentation

Image segmentation is a prevalent technique used in the field of digital image processing and analysis. It partitions an image into distinct segments or regions based on the inherent features of the pixels. This technique is used to organize pixels that are related to colour, intensity, and texture into distinct and cohesive zones, see figure 7.

JORMA International Journal of Health and Social Sciences
ISSN: 2978-0942 (Online)
Volume 3 No. 4 Issue 4 Pages 1-16

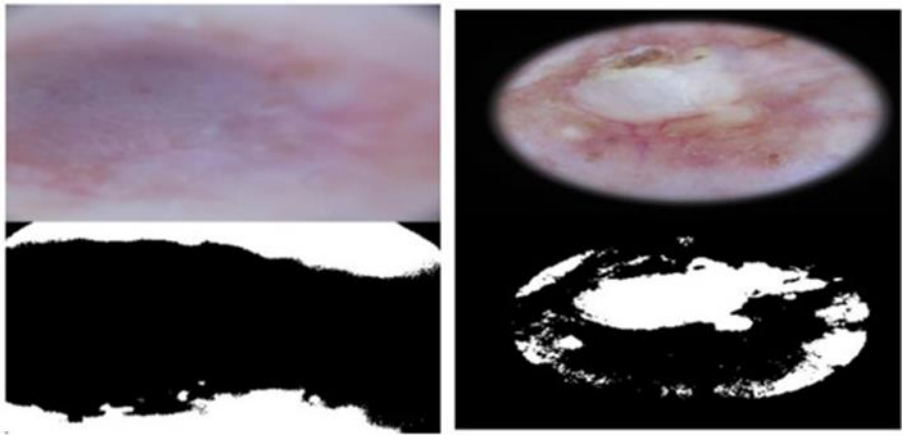

Figure 7. *segmented Images.*

Image segmentation is a process that simplifies and categorizes intricate visual data to facilitate the understanding of images by computers. It is utilized in computer vision, medical imaging, and object recognition for tasks such as object tracking, labelling, and feature extraction. Image segmentation is a crucial technique in advanced image processing and analysis. It is used to separate and describe regions of interest in a picture. It successfully divided certain categories; however, it did not accurately classify all of the images.

Data Splitting

Following the augmentation, the dataset was divided 80:20. Data splitting is a common approach used in model validation to examine the efficacy of statistical and machine learning models. This method divides a given dataset into two different subsets: training and testing. The training set is used for model training, whilst the testing set is designated for validation. This separation is critical because it eliminates any potential bias that could develop if validation were performed on the same data used for training. During model construction, the training set is used to train the model the underlying patterns and relationships in the dataset. Once trained, the model is assessed on the testing set, which it has never seen before. This approach allows us to assess how effectively the model generalizes new, previously unseen data. By comparing the model's performance on the testing set to its performance on the training set, we can determine whether it has learned the underlying patterns of the data or if it has overfitted the training data, ensuring the model's reliability and effectiveness in making predictions on actual data. According to (Sameen, Faryal, & Rahman, 2020), selecting an appropriate splitting ratio is crucial when implementing data splitting algorithms. The 80:20 ratio is a popular choice, allocating 80% of the data for training and 20% for testing.

Nevertheless, different ratios such as 70:30, 60:40, and even 50:50 is common in actual usage. It is worth mentioning that there is no universally accepted rule for establishing the best or optimal ratio for a given dataset. The 80:20 split, frequently related with the Pareto principle, is a widely used rule of thumb among practitioners. However, it is critical to understand that this ratio selection has no one- size-fits-all answer, as optimal ratios might vary based on the nature of the data, the problem being addressed, 30 and the available resources. As a result, the split ratio

should be selected after careful analysis of these factors, with the goal of providing robust model validation and generalization. So far, theoretical and numerical investigations into the finest of data splitting ratios have yielded no consensus. In 1990, Picard and Berk suggested that 25-50% of the testing should be done separately (Roshan, 2022).

Experimental Results and Analysis

Figures 8 and 9 demonstrate that the algorithms underwent training for various epochs until the training process reached its full completion.

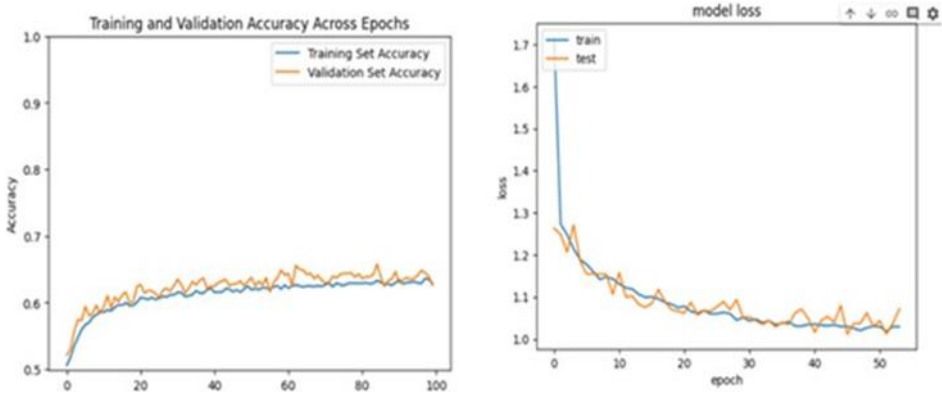

Figure 8. *Training & validation Accuracy and Model loss.*

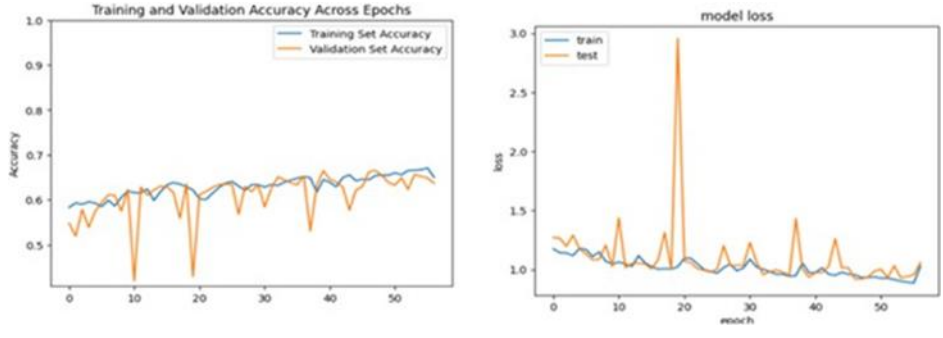

Figure 9. *Training & validation accuracy for ResNET50V2 and Model loss .*

Upon analysing the trends illustrated in figure 8, it has been concluded that both lines follow the same pattern and remain close to each other during the training phase. The earlier stopping strategy is employed to terminate the training and store the model when the lines diverge from each other. Based on the patterns, it is evident that the model is not overfit. There are three unique possibilities that will occur in this scenario: both lines will experience an increase, both lines will fluctuate together, and both lines will decline at the same time. From the zero epoch to the ninth epoch, the accuracy is growing exponentially. Subsequently, the percentage steadily rises until the 55th epoch, reaching 64%, while for the 8th epoch, it stands at 59%.

- Both Lines Rising: If both lines rise gradually and stay close, the model is learning from

the training data and generalizing well to the validation data. This suggests your model is doing well on the task.

- Both Lines Fluctuating Together: If both lines fluctuate or have similar patterns during training, the model is not overfitting but may not be working at its best. This experiment will test various hyperparameters, model topologies, and data pre-processing methods to enhance performance.
- Both Lines Decreasing: The model may need more training data or be learning poorly if both lines are decreasing. This may indicate underfitting when the model is too simplistic to capture data patterns. Here, the model should be more complicated or provide more relevant facts.

The figure 8 model loss shows that the training dataset loss decreases gradually, while the validating dataset loss fluctuates. This indicates that the model is improving during training, but the unseen dataset may result in reduced model performance due to instability. The blue line represents the training dataset loss line, indicating improved model fit, reduced mistakes, and parameter adjustments for better matches. The orange line shows the validation dataset varying epoch-by-epoch. It will happen for two reasons: The validation dataset provides more diversity or is smaller than the training dataset and the model may be overfit to specific patterns in the validation. Monitoring volatility or regularization can lessen this inaccuracy. Introducing a dropout layer or weight decay layer can successfully address validation loss instability. Hyperparameter tuning or adding data also stabilizes validation loss. Validation loss fluctuation is normal, however balancing training performance and generalization is crucial for model success in real-world applications. Figure 9 shows a contrast accuracy comparison between training and validation datasets. The blue line represents the training data set and fluctuates slightly. After 57 epochs, accuracy is 65%. In contrast, the orange line depicts the validation dataset, which exhibits a lot of fluctuation, indicating that it was not a good, trained model for future or unknown values. More data for training, parameter adjustments, or model complexity are needed to solve this problem. While the Figure9 model loss shows that the training dataset loss decreases gradually, while the validating dataset loss fluctuates. This indicates that the model is improving during training, but the unseen dataset may result in reduced model performance due to instability. The blue line represents the training dataset loss line, indicating improved model fit, reduced mistakes, and parameter adjustments for better matches. The orange line represents the validation dataset, which fluctuates epoch-by-epoch.

Model Evaluation

A test dataset is utilized in conjunction with a variety of assessment metrics during the evaluation process. This is done to gain an understanding of how well the trained model is doing. To accomplish this, the dataset is utilized, which the model does not observe. The difference between it and validation is that it verifies the accuracy of the prediction made by a trained model, whereas evaluation is a sort of examination that examines the overall performance of the model under a variety of conditions. In accordance with the various models, the following are the outcomes of the evaluation.

Table 1. Evaluation model without early stopping classification report for model.

Class	Precision	Recall	F1-Score	Support
0	0.4	0.02	0.04	88
1	0.54	0.59	0.56	337
2	0.42	0.22	0.29	260
3	1	0		27
4	0.66	0.66	0.78	1287
5	0.71	0.23	0.35	450
6	0.5	0.02	0.03	62
7	0.73	0.35	0.47	23
accuracy			0.63	2534
macro avg	0.62	0.3	0.32	2534
weighted avg	0.62	0.63	0.57	2534

In the classification report that was just presented, the performance of an eight-class model is evaluated, and the report displays precision, recall, and F1-score for each class separately as shown in the table 1. There is a 63% accuracy rate, with class 4 performing exceptionally well, while classes 0, 3, and 6 are having difficulty. There are problems that are peculiar to the class, as indicated by the macro averages of 0.62, 0.30, and 0.32 for precision, recall, and F1-score, respectively. The weighted average places an emphasis on improving minority class predictions to achieve a fair conclusion. This highlights the necessity of making improvements in the way that imbalanced data and individual class issues are handled.

Table 2. Evaluation model with early stopping classification report.

class	precision	recall	f1-score	support
0	1	0	0	88
1	0.45	0.68	0.54	337
2	0.46	0.08	0.14	260
3	1	0	0	27
4	0.67	0.94	0.78	1287
5	0.62	0.24	0.34	450
6	0.1	0.02	0.03	62
7	1	0.26	0.41	23
accuracy			0.62	2534
macro avg	0.77	0.28	0.28	2534
weighted avg	0.64	0.62	0.55	2534

Table 2 classification report evaluates a model's performance with early stopping. The accuracy

is 62%, meaning 62% of predictions are correct. Classes 0, 2, 3, and 6 struggle with low recall, indicating missed positive cases. Class 4 performs well with high precision and recall.

Table 3. Evaluation ResNet50V2 pre-trained classification report.

class	precision	recall	f1-score	support
0	1	0	0	88
1	0.26	0.2	0.23	337
2	0	0	0	260
3	1	0	0	27
4	0.73	0.37	0.49	1287
5	0.23	0.81	0.35	450
6	1	0	0	62
7	1	0	0	23
accuracy			0.36	2534
macro avg	0.65	0.17	0.13	2534
weighted avg	0.52	0.36	0.34	2534

Table 3 classification report provides an evaluation of the performance of the ResNet50V2 model. 33% is the accuracy, which measures the entire correctness of the situation. Classes 0, 2, 3, 6, and 7 have a relatively poor recall, which indicates that affirmative cases for those classes were missed. Class 5, on the other hand, has a high recall but a low precision, which indicates that it can identify many positives but also frequently generates false-positive predictions. The research emphasizes the necessity of making modifications that are better balanced to guarantee accurate projections across all classes.

Conclusion

Skin cancer is a significant health concern. Timely detection and intervention enhance the likelihood of a successful recuperation, whereas extended exposure poses a risk to one's well-being. A contrast analysis is performed between a Sequential Convolutional Neural Network and ResNet50V2, which is a pre-trained transfer learning model. Photos that are not augmented cause the model to overfit, hence hindering its ability to work with future values. In this study, a selection of 80 × 20 pictures was made due to the need for separate training when dealing with a large dataset. The accuracy of the sequential model is approximately 64%, and it consistently maintains validation loss and accuracy that align with the training data. It will produce consistent and satisfactory outcomes with an accurate rate of 64%. The ResNet50V2 model achieves an accuracy of approximately 63%. However, its validation loss and accuracy exhibit fluctuations throughout training, suggesting the presence of unknown instabilities. To overcome this, it is necessary to augment the dataset and raise the size of the set for transfer learning. Transfer learning is effective for specific datasets that satisfy the prerequisites of pre-trained models. Furthermore, artificial intelligence plays a vital role in promptly detecting skin cancer, thereby safeguarding a substantial number of individuals from this devastating illness.

Research Limitations

One other issue is that segmentation produces false results for some of the images that are chosen at random. This is a major problem with segmentation. The primary aim of this project is to analyses each image to determine whether it has been properly segmented. As a result, the segmentation technique is not feasible for this project due to the time constraints. As the processing power is also a crucial factor in the training process, it plays a crucial role in efficient time management.

This research makes use of a system that is based on graphics processing units (GPUs) to save time. When compared to the typical GPU-based systems, the TPU-based system, the artificial intelligence-based system, such as the jetson nano board, and the supercomputer, which have previously been presented, are operating at a significantly quicker rate. Additionally, the issue of classed imbalance is a concern because it will lead to overfitting during the training process.

Acknowledgments

Funding Statement: This research was not funded by any organisation.

Competing Interests: None

Data Availability Statement: The data, code and other materials will be available upon request.

Ethical Standards: The ethical considerations that are related to the application of AI and ML techniques in healthcare are acknowledged in a cursory manner in the study; nevertheless, the paper does not dive into these considerations in depth. In the future, research should investigate the ethical consequences of deploying these models, including concerns regarding the privacy of patients' data, the consent of patients, and the biases that are introduced by algorithms.

Author Contributions: Conceptualization: R.I; Methodology: R.I Data curation: R.I; M.M Data visualization: R.I; D.M; M; M.M Writing original draft: R.I; D.M. Proof Reading: D.M; All authors approved the final submitted draft.

References

Aima, A., & Sharma, A. K. (2019). Predictive approach for melanoma skin Cancer detection using CNN. In *Proceedings of international conference on sustainable computing in science, technology and management (SUSCOM), Amity University Rajasthan, Jaipur-India*. DOI: http://dx.doi.org/10.2139/ssrn.3352407

Alavi, M. (1994). Computer-Mediated Collaborative Learning: An Empirical Evaluation. *MIS Quarterly*, *18*(2), 159–174. https://doi.org/10.2307/249763

Codella, N. C., Gutman, D., Celebi, M. E., Helba, B., Marchetti, M. A., Dusza, S. W., ... & Halpern, A. (2018, April). 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). In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)* (pp. 168-172). IEEE. DOI: https://doi.org/10.48550/arXiv.1710.05006

Du-Harpur, X., Watt, F. M., Luscombe, N. M., & Lynch, M. D. (2020). What is AI? Applications of artificial intelligence to dermatology. *British Journal of Dermatology*, *183*(3), 423-430. DOI: 10.1111/bjd.18880

Haenssle, H. A., Fink, C., Schneiderbauer, R., Toberer, F., Buhl, T., Blum, A., ... & Zalaudek, I. (2018). Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Annals of oncology*, *29*(8), 1836-1842. PMID: 29846502. DOI: https://doi.org/10.1093/annonc/mdy166

Hassani, H., Silva, E. S., Unger, S., TajMazinani, M., & Mac Feely, S. (2020). Artificial intelligence (AI) or intelligence augmentation (IA): What is the future?. *AI*, *1*(2), 8. https://doi.org/10.3390/ai1020008

He, K., Zhang, X., Ren, S. and Sun, J. (2016). Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, 27-30 June 2016, 770-778. https://doi.org/10.1109/CVPR.2016.90

Introna **L** (1997). *Management, Information and Power*: A narrative of the manager involved. London: MacMillan (PDF) The History of Information: Lessons for Information Management

Joseph, V. R. (2022). Optimal ratio for data splitting. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, *15*(4), 531-538. DOI: https://doi.org/10.1002/sam.11583

Nawaz, M., Mehmood, Z., Nazir, T., Naqvi, R. A., Rehman, A., Iqbal, M., & Saba, T. (2022). Skin cancer detection from dermoscopic images using deep learning and fuzzy k-means clustering. *Microscopy research and technique*, *85*(1), 339-351. DOI:10.1002/jemt.23908

Park, Y. J., Kwon, G. H., Kim, J. O., Kim, N. K., Ryu, W. S., & Lee, K. S. (2020). A retrospective study of changes in skin cancer characteristics over 11 years. *Archives of craniofacial surgery*, *21*(2), 87. DOI: https://doi.org/10.7181/acfs.2020.00024

Poonkuzhali S., Anu Barathi B.U., & Vinodhkumar S. (2022). *Transfer learning approach for diagnosing skin cancer with deep convolutional neural network*. Information and Communication Technology for Competitive Strategies. DOI:10.21786/bbrc/13.11/13

Sameen, D., Faryal, I., & Rahman, H. (2020). Skin cancer disease detection using image processing techniques. *LC International Journal of STEM*, *1*(3), 50-58. https://doi.org/10.5281/ZENODO.5148277