

Skin Lesion Classification Based on Deep Convolutional Neural Networks Architectures

Ayesha Athar¹, Eman Shahid¹, Maheen Khalid¹, Dr. Muhammad Munwar Iqbal¹

21-CS-22@students.uettaxila.edu.pk 21-CS-13@students.uettaxila.edu.pk 21-CS-07@students.uettaxila.edu.pk munwar.iq@uettaxila.edu.pk

¹Department of Computer Science, University of Engineering and Technology, Taxila, Punjab, Pakistan

ABSTRACT

Melanoma or skin cancer is one of the most frequent types of cancer and can be caused by numerous skin diseases which are known to be divided into several categories regarding the morphological, color, texture, and structure characteristics. Therefore, key factors that have helped the decrease in mortality rate include early and timely detection of the appearance of cancer cells. Nevertheless, the sizes of dermoscopic images, which play a decisive role in diagnostics, may be significantly smaller due to shadows, artifacts, and noise resulting from the use of modern imaging techniques. Some of these effects include the following; They are as follows These aspects can hamper the possibility of detecting the cancerous tissues in the body correctly. To resolve all these challenges, deep-learning neural networks have been used in order to improve the quality and fluency needed in skin image cancer analysis. This is due to the fact that, through deep learning techniques, results in skin cancer image registration and processing that leads to better and early diagnosis can be enhanced. This paper seeks to discuss the work done on the topic of using deep learning on skin cancer diagnosis and classification to enhance medical excellence. Another approach is the use of dermoscopy and the integration of sophisticated imaging with the neural network to overcome the barriers of dermoscopic imaging. The usage of such a strategy also helps to improve the accuracy and time taken to diagnose skin cancer, which helps in improving the quality of patient care and their likely outcome. Applying the concept of deep learning, tremendous advancements in the diagnosis and treatment of skin cancer may be achieved throughout the enhancement of the rate of survival among the affected individuals.

INTRODUCTION

Cancer is not just one disease but a collection of maladies where cells grow abnormally and invade surrounding tissues. Any cancer can be deadly, but some are more common than others: skin cancer tops the list as the most prevalent form of this deadly scourge—split into melanoma or non-melanoma types. Malignant lesions cost a lot in healthcare bills and take many lives; thus, researchers are working on algorithms that can detect early melanomas with high accuracy and flexibility. Finding out whether these malignant cells have spread often involves expensive tests like ELM; however, if detected early enough, further metastasis can be prevented using low-cost procedures.

This method involves using dermatoscopy to enhance the visualization of patterns by utilizing a lens with a source to identify colors, textures, pigmented tissue, structures, effects and more. Dermatologists identify melanoma by observing characteristics, like shape, border irregularities, color variations, size differences and changes over time (referred to as the ABCD rule). However, the limitations of dermoscopic imaging caution. While techniques for image processing have been developed to enhance Computer Aided Detection (CAD) systems and strategies for segmenting and classifying Pigmented Skin Lesions (PSL) thereby aiding in the diagnosis of patients without traumatic medical procedures. Progress in machine learning in neural networks has resulted in breakthroughs across various domains. Deep neural systems can play a role in processing images such as skin lesions within the medical field. Symptoms like skin lesions abnormalities in skin texture or coloration, pain sensations or granulomas are signs. Healthcare providers can utilize machine learning methods to identify and categorize skin lesions from images before making decisions impacting a patient's well-being. The process of detecting and diagnosing melanoma cancer typically involves four stages; image preprocessing, segmentation of lesions from images, feature extraction, from these segmented areas followed by lesion classification. Litjens et al. In 2017, image classification, object detection, segmentation, registration, etc. The authoritative concept of deep learning related to medical image analysis has been reviewed, for example, in 2018, Okur et al. Research in melanoma and dermoscopic imaging focuses on the techniques used to present and detect melanoma. In 2019, Cassani et al. A comparative study of recent deep learning approaches in dermoscopic imaging for skin lesion classification. In another study conducted in the same year, Munir et al. (2019) demonstrated their chronological success in using machine images in cancer diagnosis and medical imaging. In this article, the authors aim to review the current deep learning architecture with key concepts and challenges for cancer diagnosis in dermoscopic images.

Skin most cancers is the result of exposure to UV rays, which harm the DNA in skin cells. This damage changes genes, inflicting ordinary growth and swelling of skin cells. Genetic defects reason pores and skin most cancers, pores and skin lesions, allergic reactions, and most cancers cells. The destruction of most cancers cells is an critical component in this system. In excessive instances, cancer symptoms can cause dying. Melanoma a kind of pores and skin cancer, has an 8% mortality charge, that is taken into consideration excessive. Skin most cancers is classified into melanoma pores and skin cancer (MSC) and non-cancer pores and skin cancer (NMSC).

The ABCD code has been used for the clinical diagnosis of visceral and skin cancers as follows:

- A: Individual irregularities: Bilateral examination of pores and skin lesions for aspects, shape and color.
- B: Border function: Classified according to the edges of skin lesions as either well defined and smooth. Otherwise, if the lines are jagged, rough, and irregular, the lesions are probably melanoma.
- C: Distinctive color: Melanoma appears in dark, red, brown and tan patches on one skin layer.

- D: Extent of symptoms: Skin lesions larger than 6 mm are usually indicative of cancer.

The aims and objectives of this research includes:

1. To provide new dataset sources for deep learning model training and improve the efficiency and precision of skin cancer detection and diagnosis.
2. To establish a foundation for building a highly accurate deep learning based system to detect skin cancer and classify them automatically.
3. To study the role and efficacy of transfer learning and pre-trained models in skin lesion classification.
4. To evaluate and compare various deep learning structures in clinical skin lesion classification.
5. To discuss valuable suggestions and review opportunities for future research related to the enhancement of skin cancer detection models.

In this paper, DenseNet and Convolutional Neural Networks (CNN) models have been applied to automate the process of skin cancer detection and classification using HAM10000 dataset. It is mainly concerned with understanding different preprocessing techniques like resizing, image augmentation, and normalization of the input images so as to obtain more quantity and better quality of training data. Having trained and optimized DenseNet we experience the improvement of the classification accuracy. The results of the model which are demonstrated in accuracy, precision, recall, and F1-score show how well it works for melanoma and skin lesions diagnosis. This study not only presented how the DenseNet and CNN architectures can be used for medical image analysis but also compared and contrasted their performance with traditional CNN models. These results highlight the potential of deep learning for early skin cancer diagnosis and provide direction for future studies to fine-tune and rigorously assess these models for in-clinic use.

The remainder of this paper is organized as follows:

- II. Literature Review**
- III. Research Methodology**
- IV. Results and Discussions**
- V. Conclusion and Future Work**
- VI. Citations and References**

LITERATURE REVIEW

Although noticeable progression in the diagnosis and classification of skin cancers has been noted, the use of deep learning approaches has embraced dermatology as a specialty. Initial attempts to apply dermatoscopic imaging for distinguishing melanomas from neoplasms was not very efficient because The datasets which were available at that time were relatively small and heterogeneous and hence, did not augur well for the formulation of efficient automatic diagnosing systems. To overcome these limitations HAM10000 was formulated that has got more than 10000 dermatoscopic images collected from various population. It is now considered as fundamental dataset in academic machine learning and can be found in the ISIC archive along with the contextual information, it therefore offers a sound base for training and testing neural network. This dataset has been used by recent research works in training and enhancing deep learning models through CNN and DenseNet models whose results were observed to increase skin cancer detection accuracy and reliability. These progresses have shown the fact that deep learning can be a game changer in the field of dermatological diseases, resulting in early diagnosis and improved prognosis for patients.

In a radical inquiry, Asad and company were involved in an interesting act of recognizing handwritten numbers. They brought the most popular MNIST dataset to life using convolutional neural networks (CNN). They did this by investigating how hand written numerals change with time—and then used CNNs designed

for visual image analysis which are good at detecting changes over time in images so that their models could become more accurate. This work is important because apart from showing that we understand the techniques involved in CNNs, it also advances them beyond what we currently have; thus representing a real contribution towards deep learning.(Shetty et al., 2022)

In this paper, they employed many kinds of neural networks which were able to create a classification system better than any other in the world. They used various networks such as CNNs in order to increase overall performance and reduce false positives. Their progress is amazing, but lets not forget about the issues that come with more complex things and needing more resources. This way works best when we all bring something different to the table so let's keep working on it. It's clear that we have a long way to go in terms of finding the right balance between how fast can computer do things and how accurate diagnosis is its important for doing further research towards making this better for patients.(Popescu et al., 2022)

In this seminal work by LeCun et al., one sees that the convolutional neural networks are very good at accurately classifying handwritten digits in the widely-used MNIST dataset. LeCun and colleagues have gone beyond all previous attempts to classify digits in a way that far surpasses previous efforts. The key here is the application of computer vision techniques—specifically, CNNs—that allow scanning an image layer by layer, much like a visual cortex. Perhaps the most tantalizing aspect of this study is that these techniques seem to allow the representation of digits so highly that they even seem capable of recognizing handwriting as well as a human on a controlled MNIST sample. However, the authors state that this certainly does not mean their approach should be considered at all satisfactory for such more varied, complex datasets as the real world. The major challenge that remains is to extend this work beyond the ultra-controlled environment of the MNIST dataset. Building on this work will be the major challenge that remains, with the impact of this promising research being all too evident in the future work.(Lopez et al., 2017)

A linchpin paper came out of the opulent chambers of (Garg et al., 2019) presenting a new Decision Support System (DSS) that should change forever the way skin cancer is both diagnosed and staged. The study took advantage of Convolutional Neural Networks (CNN) a state-of-the-art computing design that is advancing quite quickly in this field. This combination of a group of sincerely enthusiastic machine learning and medical imaging experts brought to life a CNN-based model that is second to none when it comes to recognizing a plethora of appearances of skin lesions. The study is a meticulous and beautiful meta-painting of what deep learning can do to improve the accuracy of the diagnosis but also what type of advice it can give in clinical matters. Surprisingly, though, the article published is the least great of this study. There are multiple key points which, aside from obvious good side of the story, would require much longer extolling the disadvantages. One of the darkest shadows comes from the fact that they used very, very limited and also very, very, very poor data as their database material. A very bad omen, as bad as any, telling us robustly hungry algorithms can be really fed without any problems, the ones which are able to grow up healthy.

Albahar in survey research work, in 2019, describes the use of Convolutional Neural Networks (CNNs) for skin lesion classification. He presents a novel way to regularize our models that would make them better. Although the study does not mention explicitly about the datasets used, it is assumed that the ISIC datasets have been used. The paper aims at early detection of cancer in skin lesions using the classification mentioned before. The use of CNNs results in the ability of generalization and reduction in overfitting. If the regularization used is correct, the accuracy should be increased. The paper presents related challenges like e.g. the quality and diversification of online learning. Similarly, it points out that probability of success depends on many factors and it cannot be transferred to recreation activity or hospital context. Although the authors recommend that research be continued to make the model reliable in other populations and settings, the research might also show irrelevance in other populations and settings.(Albahar, 2019)

"Skin Cancer Detection Using Machine Learning Algorithms" is an accurate article of an author which focuses on the detection of skin cancer by a method of super smart machine learning algorithms. The task of computer vision for skin cancer detection is accomplished with the help of CNNs and datasets like ISIC and HAM10000, which attempt to demonstrate the findability of deep machine learning models such as Random Forest, SVM, CNN, and DenseNet. The models are compared in great detail that the set of datasets shows how important a transfer learning feature can be and how much a pre-trained model can contribute to a higher accuracy of detection. It can be deduced that for CNN and DenseNet architectures accuracy, precision, recall, and F1 scores have been higher. However, the approach of this paper to class imbalance and the choice of machine learning algorithms could still be elaborated on, such as using other models and techniques that help mitigate the limitations of the datasets and increase the efficacy of these models in classifying the data. (Ramachandro et al., 2021)

The paper "Skin Cancer Detection and Classification using Deep Learning" deals with melanoma diagnosis through the introduction of deep learning methods that derive results on the TensorFlow backend and Keras libraries. In the model, HAM-10000 dataset which includes over 10,000 dermoscopic images is used and preprocessing step is introduced as a technique to convert the images to grayscale and filter them. This research report depicts the application of the MobileNet optimized method for classification, which parallelly achieved noteworthy accuracy improvements over the ones that were existing in the past, with approximated Top-1 accuracy of 89-90%. So the conducted work is the reflection of how deep learning can change diagnosis procedure for skin cancer, and this aspect needs additional research to improve classifier optimization and data exploration for reliable and high-accuracy classification. The technique brings out the usefulness of deploying state-of-the-art machine learning algorithms along with large datasets in medical imaging. However, the results also highlight the need for more research, especially on imbalanced datasets and generalizability of the algorithms to give a good diagnosis for a wide range of patient populations. (Said et al., 2022)

This research project explores the breakthroughs that came from deep learning techniques about skin cancer diagnostics. The team has used the Deep Convolutional Neural Network, DCNN, model in tandem with the commonly used VGG16 and VGG19 models. Through the training and testing these models with the HAM10000 dermatoscopic image dataset, an asset of the studies of dermatological kind, the team seeks to give better accuracy and efficiency for the classification of skin cancer in better outcomes for patients. This research looks to be a breakthrough, from learning techniques for a solution to these kinds of interesting problems. The state-of-the-art model, the newly proposed DCNN architecture of construction from scratch, confirms remarkable classification accuracy and overcomes the challenge of the dataset imbalance in a very high-level manner using sophisticated techniques of image augmentation. The comparison of the state-of-the-art model with the VGG models shows that the newly proposed DCNN is ahead in the accuracy and loss metrics. This supports that the model will likely significantly enhance skin cancer diagnostic processes. The strong recommendation to apply these models in a range of datasets and to incorporate additional patient metadata in an effort to further improve accuracy of diagnosis becomes important in a range of dermatological artificial intelligence research. It points to the potential of deep learning techniques for enhancing diagnostic capability in the research of medical work. (Garg et al., 2021)

The paper elaborated in the intriguing contents above compares the accuracy of Support Vector Machine and Convolutional Neural Network in classifying and diagnosing the existence of skin cancer. The classification is achieved through the use of trained computer algorithms to explore the ability of a classifier to determine the class and label of data items. This can be summarized as an algorithm that begins as a decision-making program until it's trained to execute and deduce logic in a sensible way. The target audience is people

interested in training a skin cancer classification model or making the existing models more efficient.(Pavithra & Geetha, 2023)

The paper discusses this unique method by employing SVM alongside image processing algorithms for timely skin cancer diagnosis. The article explains how noise reduction, image enhancement and feature extraction by GLS helps to differentiate between cancerous and non-cancerous skin lesions. This research significantly influences the dermatology profession by providing the noninvasive, accurate and fast diagnostic tool. However, it indicates that for a more precise diagnostic process and applicability, a thorough investigation of many more intricate algorithms and extensive datasets is essential. The next step is therefore designing a framework that will involve the integration of various machine learning models and more advanced image processing methods in order to develop more sophisticated diagnosis process for the skin cancer.(Ansari & Sarode, 2017)

This paper introduces a novel approach to the classification of benign and malignant skin lesions into the radial graph embedding feature space. It leverages deep features derived from well-established CNN architectures, namely AlexNet, VGG16, and ResNet-18 to train SVM classifiers that are then hybridly fused. This proposed method is applied to the ISIC 2017 dataset with impressive performance: an AUC of 83.83% for melanoma and 97.55% for seborrheic keratosis. The proposed method is a hybrid deep learning and machine learning approach with high-value potential to improve the accuracy of skin lesion diagnosis so that patients at risk for skin cancers may get diagnosed faster and then treated faster, thus contributing to better outcomes. High AUC scores show the promising potential for supplementing the work of dermatologists. (Mahbod et al., 2019)

Faouzi Adjem et al. introduce one of the advanced approaches to distinguish melanoma from non-melanoma skin cancers. In this study, Local Binary Patterns (LBP) are used to extract the textural information and are used as classification features with a Support Vector Machine. They achieved the accuracy of 76.1%, with 75.6% sensitivity and 76.7% specificity. This demonstrates the critical role of early and accurate skin cancer classification for the process of diagnosis and planning of treatment. Added to this, LBP along with SVM aids in the achievement of automated systems for detection. This approach, together with the achieved accuracy, emphasizes early recognition and is thus of importance in dermatology. This study shows that it sets a very interesting goal in terms of early detection and very high accuracy achieved, and will be of importance for diagnosis and treatment planning of skin cancers. (Kassem et al., 2021)

The paper presents an integrated diagnostic framework that incorporates deep convolutional networks in the processes of segmentation of skin lesions and for their classification. Deep FrCN is utilized for segmentation, and it involves four classifiers: Inception-v3, ResNet-50, Inception-ResNet-v2, and DenseNet-201. The proposed method has been applied using ISIC datasets available for years 2016 to 2018. The approach has shown high accuracy in the range of 81.79% in the classification. The method increases its accuracy when it works at the cost of the data augmentation and balancing. However, the study has concluded that it was not effective in handling the imbalanced data and it needed a lot of datasets so that the model generalizes more.(Al-Masni et al., 2020)

In this study, an innovative approach utilizing image processing and machine learning is presented for the purpose of skin cancer classification especially melanoma detection. The dataset adopted was ISIC-ISBI 2016 of which diverse techniques like contrast stretching, segmentation using OTSU method and feature extraction using GLCM, HOG etc., PCA for dimensionality reduction were included in the proposed model. We have used Quadratic Discriminant, SVM with Medium Gaussian, and Random Forest techniques, in which Random Forest technique observed the highest accuracy of 93.89%. The research has succeeded in

solving the class imbalance problem by means of SMOTE sampling and a fruitful feature selection method based on the wrapper approach. (Javaid et al., 2021)

This work uses a VGG-16 model from TensorFlow to classify skin diseases from a dataset of 16,453 clinical images from Aarhus University Hospital, Denmark. The goal is binary classification of different tasks, such as psoriasis vs. eczema, acne vs. rosacea, and CTCL vs. eczema; it also aims at 77%-plus accuracy for the best model, VGG-16P. Among the limitations are some small datasets for some conditions and possible racial bias in the preponderant Fitzpatrick skin types II and III. (Thomsen et al., 2020)

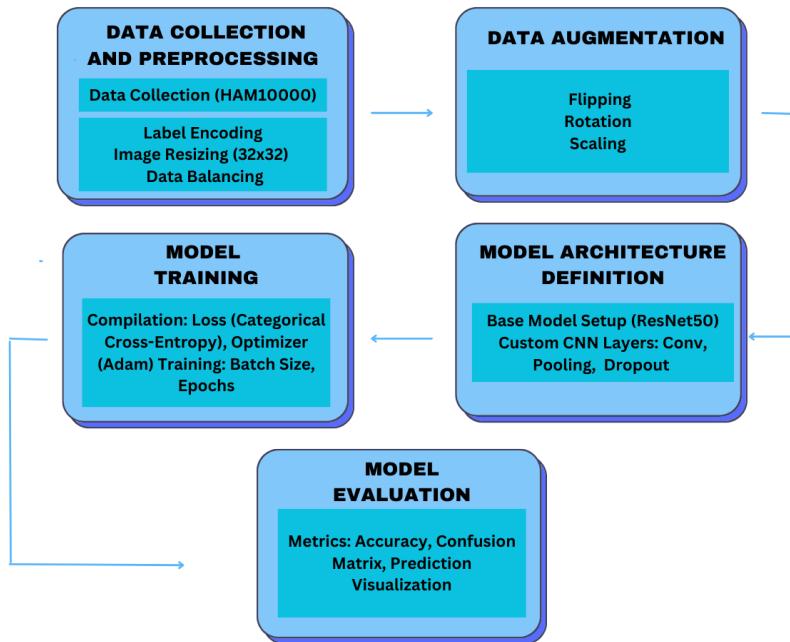
Table 1: LITERATURE SUMMARY

Sr. No.	Author	Problem	Dataset	Technique	Performance	Limitation
1.	Roshan Fernandes et al., 2022	Classification of skin lesion images	HAM10000 dataset	CNN, ML	95.18%	Overfitting, Data Dependency
2.	Dan Popescu, Mohamed El-khatib and Loretta Ichim, 2022	Classification of skin lesion images	HAM10000 dataset	CNN	86.71%	Optimization, Diversification
3.	Adria Romero et al. 2017	Classification of skin lesion images	ISBI 2016 Challenge dataset	CNN, VGGNet	79.74%	Scalability, Generalization
4.	Rishu Garg, Saumil Maheshwari and Anupam Shukla, 2019	Classification of skin lesion images	MNIST HAM-10000	CNN	90.51%	Prediction of result classification
5.	Marwan Ali Albahar, 2019	Classification of skin lesion images	ISIC archives	CNN	97.49%,	Cannot be used for feature selection or feature reduction, Choosing a suitable value of λ is difficult
6.	M.Ramachandro et al., 2021	Classification of skin lesion images	ISIC, MNIST: HAM10000	CNN	94% on ISIC, 95% on HAM10000	Class imbalance issue, Model overfitting
7.	Raed A. Said et al., 2022	Detection and classification of melanoma	HAM-10000	CNN	Top-1 accuracy of 90% Top-1 precision of 90%	Low classifier learning rates, not represent all variations of skin cancer globally.
8.	Nour Aburaed et al., 2020	Early skin cancer detection and classification	(HAM) 10000	VGG16, VGG19, DCNN	99%	Dataset imbalance, no utilization of metadata

9.	A. Pavithra and B. T. Geetha, 2023	Prediction and detection of skin cancer	20 samples from a predefined skin cancer classification set to compare the accuracy of SVM and CNN	SVM and CNN	CNN: an accuracy of 95.91%, SVM: an accuracy of 94.30%.	The relatively small sample size of 20 means that this study has low generalizability and diagnostic performance.
10.	Uzma Bano Ansari and Tanuja Sarode , 2017	Early and efficient skin cancer detection	Skin cancer dermoscopy	SVM, GLCM	accuracy of 95%	Lack of detail on the dataset's size and diversity
11.	Amirreza Mahbod, Gerald Schaefer, Chunliang Wang, Rupert Ecker , Isabella Ellinger, 2019	Classification of skin lesion images	ISIC 2017	CNN architectures—AlexNet, VGG16, and ResNet-18	Melanoma: 83.83% seborrheic keratosis: 97.55%	Overfitting, Data Dependency
12.	Faouzi Adjed1 & Gardez, 2021	Classification of Skin Cancer	ISIC archives	SVM and LBP	76.1%, with sensitivity and specificity rates of 75.6% and 76.7%	Low classifier learning rates, doesn't represent all variations of skin
13.	Mohammed A. Al-masni, 2020	skin lesion classification segmentation using CNN	ISIC datasets from 2016 to 2018	Inception-v3, ResNet-50, Inception-ResNet-v2, and DenseNet-201	81.79%	Handling imbalanced data and the need for extensive datasets to improve the model's generalization
14.	Javaid et al. 2021	image processing and machine learning based skin cancer classification	ISIC-ISBI 2016	Quadratic Discriminant, SVM (Medium Gaussian), and Random Forest	93.89%	Not explicitly discussed

15.	Kenneth Thomsen1 & Winther2, 2020	Classification of Multiple-Lesion using deep-learning	16,453 clinical images from Aarhus University Hospital.	CNN	77%	Potential overfitting due to small datasets for certain conditions and racial bias due to the predominance of Fitzpatrick skin types II and III
-----	-----------------------------------	---	---	-----	-----	---

RESEARCH METHODOLOGY



The limited number and high variety of current dermatologic images makes it difficult to train neurons to automatically detect pigmented skin lesions. We extract the HAM10000 (Anti-human equipment and 10000 training images) database to address this issue. We collected dermatoscopic images from different subjects that were captured and stored by different methods. Because of this diversity, we have to develop a semi-functional workflow with specially trained neurons and use different acquisition and purification methods. The final dataset consists of 10,015 skin care images prepared for academic machine learning and freely available in the ISIC archive. Pathology confirms more than 50% of the lesions, and in other cases is accurate. or confirmation by in-vivo confocal microscopy, expert consensus, or follow-up.(see Table 2)

Table 2: THE HAM10000 DATASET PAGE

Design Type(s)	<ul style="list-style-type: none"> the purpose of creating a database objectives of data integration image processing
-----------------------	--

	<ul style="list-style-type: none"> transformation objectives
Scale Type	<ul style="list-style-type: none"> Skin wounds
Type of Technology	<ul style="list-style-type: none"> Digital treatment
Type of Factors	<ul style="list-style-type: none"> Assessment Diagnostic method Age Gender
Example feature(s)	<ul style="list-style-type: none"> Homo sapiens body skin

On bare inspection, dermatoscopy finds pigmented skin lesions. In order to train artificial neural systems for automatic lesion diagnosis, it is also crucial. Earlier initiatives like Binder et al. In 1994, the small sample size and focus on the fewest types of lesions limited the potential of employing dermatoscopic imaging to distinguish melanomas from neoplasms. On the other hand, expectations for the creation of sophisticated autonomous diagnostic systems have grown in light of recent developments in machine learning.

A significant number of annotated photos are needed to develop a neural-based diagnostic algorithm, however the quantity of credible diagnoses for high-quality dermatoscopic images is restricted to lesser disorders.

Dermatoscopy detects pigmented skin lesions on naked examination. It also plays an important role in training artificial nervous systems for automatic lesion diagnosis. Early efforts such as Binder et al. In 1994, the promise of using dermatoscopic imaging to differentiate melanomas from neoplasms was limited by the small sample size and attention to the fewest types of lesions. However, recent technological advances and machine learning have increased expectations for the development of advanced automatic diagnostic systems.

Developing accurate diagnostic algorithms through machine learning requires large annotated datasets. However, in dermatology, publicly available high-quality image datasets have historically been limited in scope, often focusing on only a few diseases.

Models based on these early efforts did show promise, but were limited by the narrow scope of their training data. The 2010 PH2 dataset featured 200 pictures of melanocytic lesions, a prerequisite first step, but did not capture even the full range likely to be seen in a single clinical cohort. Even if useful as a broad exploratory exercise, it was not surprising that any algorithms constructed exclusively from such a limited set of examples at that stage performed no better than moderately on true clinical samples. To produce generalisable models, and thus functional assistants, real diversity of inputs is required. The 2018 HAM10000 , which publicly provided 10,000 annotated dermatoscopic photos, is a welcome step in the right direction. Not only will it allow far more effective algorithm development, but it will also provide an easy opportunity for previously impossible direct comparisons between machine and human performance.

The dataset's thoroughness and well-documented information help us make meaningful progress. In machine learning, having large, well-labeled datasets is crucial, and the extensive scope and accessibility of HAM10000 now make it easier for us to conduct more impactful research on automated dermatological diagnosis. A new method of data sharing that is a collaboration of its own is needed for achieving the unattainable and — as a result — advancing care for patients. We must make sure we do not limit ourselves in resources, but we continue broadening and varying them to bridge the distance between research and practical clinical reality.

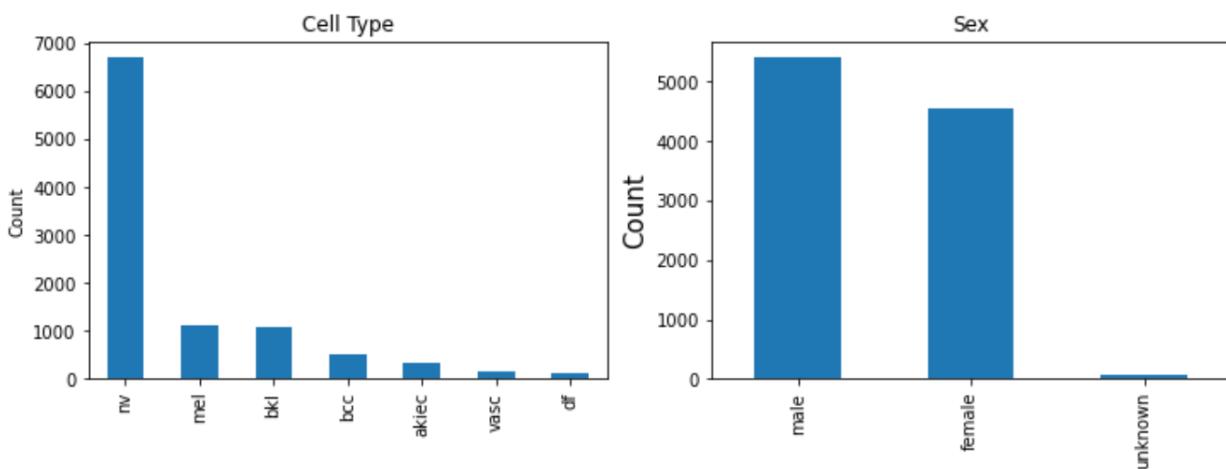


Figure 2: HAM10000 Dataset Insights on the basis of Cell types and Gender

Figure 2 demonstrates two subplots that deliver key insights into which cells are present and the division between males and females within the dataset of this data. The first subplot highlighted in the upper left and corresponding to the bar plot is dedicated to the number of cell types (dx). The bar chart displays the various types of cells in the form of stacked bar graph, with the height of each bar being the frequency of occurrence of the mentioned cell. This electronic illustration gives the reader a good perception of how the dataset is composed of distinct cell types.

In the upper-right quarter, the second-story line graph embodies the same distribution and sex (X) values. Each segment in this plot represents a sex category, which height visualizes how often this phenomenon has occurred in that specific area. This image makes a comparison at fault of the frequencies of male and female categories very easy. The combination of the two subplots into the same dropdown facilitates viewers' ability to graphically contrast the cell population and sex. The labels placed on titles and y-axis along with these plots help us in interpretation of data lodged therein, thus favoring enhanced understanding of distribution.

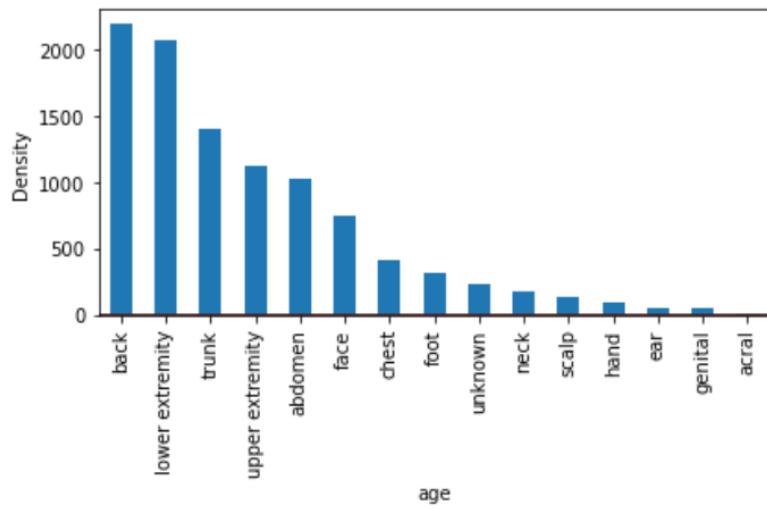


Figure 3: HAM10000 Overview of Localization Distribution

In Figure 3 the subplot, gives an overview of localizations distribution with a bar plot. Each bar represents a certain category of a localization; the height represents the number of occurrences of that localization in the dataset. This visualization provides insights into the frequency of different localization categories present in the data.

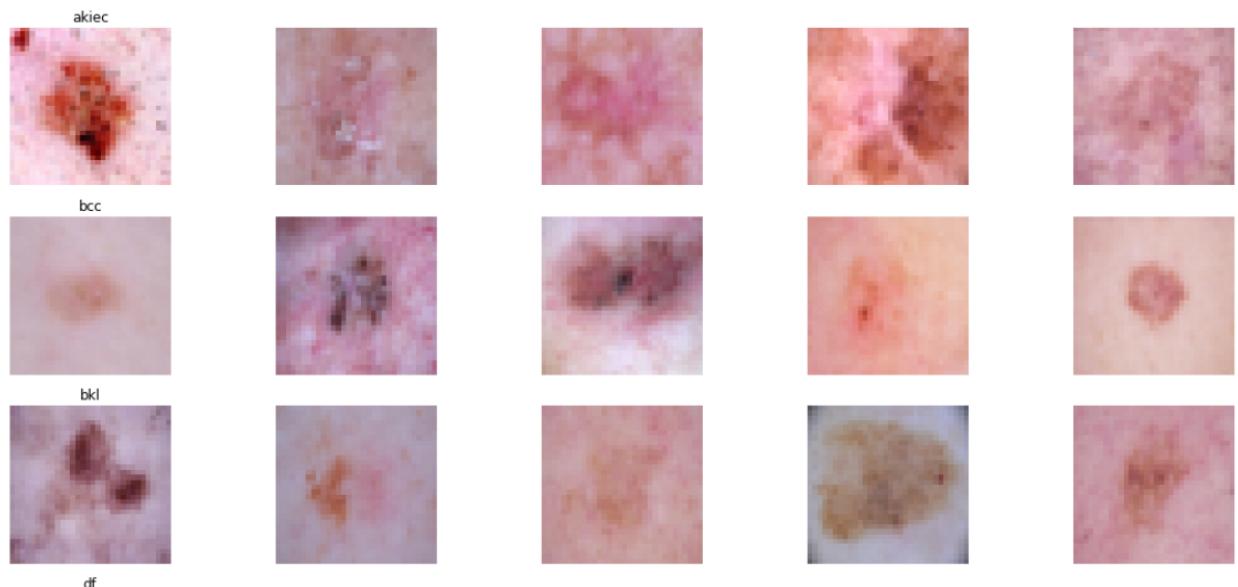


Figure 4: A subset of images associated with various skin condition categories (a)

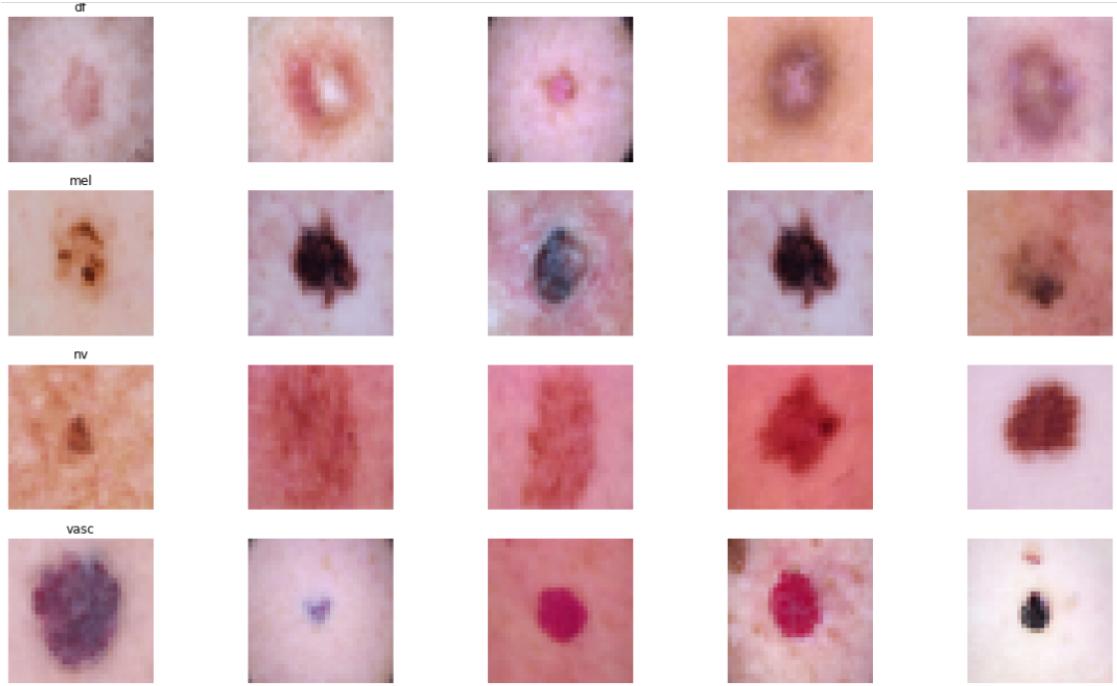


Figure 5: A subset of images associated with various skin condition categories (b)

Figure 4 and Figure 5 present a two-dimensional grid, where every row covers one of the skin problem groups and every column depicts a specific illustration representing that category (see Figures 5 and 6). The n_samples variable , which is set to 5 in this case , is acted as a sample number. The program successively isolated each row and used a random selection of appropriate images of each skin condition that were then placed in separate subplots. The images have been retrieved from an DF_skin_balanced which may have skin's associated information such as image data and other related information. This visualization will be achieved by presenting the users with an overview of images made to represent distinct skin diseases they can localize, compare and understand in an orderly way.

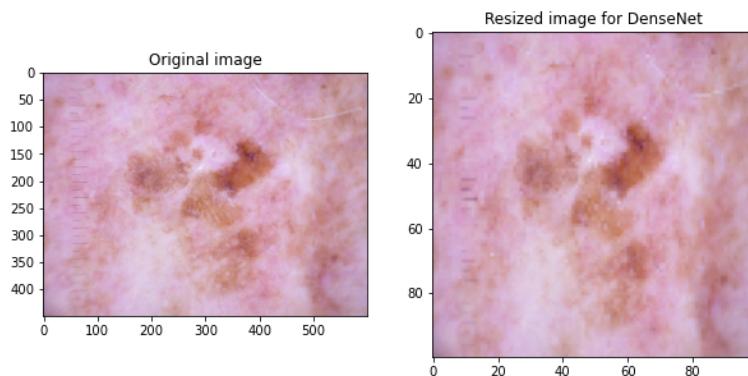


Figure 6: Original Vs. Resized Image

In the preprocessing stage for the skin lesion classification models, categorical labels are converted into numerical values using scikit-learn's `LabelEncoder` to facilitate machine learning algorithms. To address the issue of class imbalance, the dataset undergoes resampling of each class using the `resample` function from scikit-learn, ensuring equal representation across all classes. Additionally, images are uniformly resized to 32x32 pixels, standardizing the input dimensions for the Convolutional Neural Network (CNN) model. This resizing step is crucial for maintaining consistency and optimizing the model's performance.

displays the 1st image from the specified directory, first loading it with OpenCV (cv2), then resizing to a little/small size for processing with DenseNet which is deep learning model. The original and the resized image are demonstrated in a single frame which makes comparison effortless and easy.

For data augmentation, the dataset is enhanced by generating additional images through resizing, effectively increasing the size of the training set. This augmentation technique not only expands the dataset but also improves the model's robustness and generalizability. By augmenting the dataset, the model becomes more reliable and resilient, capable of better performance when applied to real-world scenarios. These preprocessing and augmentation steps are integral to preparing the dataset, ensuring that the models are trained on a well-balanced and sufficiently large dataset to achieve high accuracy and reliability in skin lesion classification.

Figure 7 shows the function generate_new_image(img2) which rotates and flips the original image(img2) with OpenCV image processing methods. It does this by feeding the function into the programming language that will be used to create the whole bunch of new images, arranging them next to the original picture in a grid visualization.

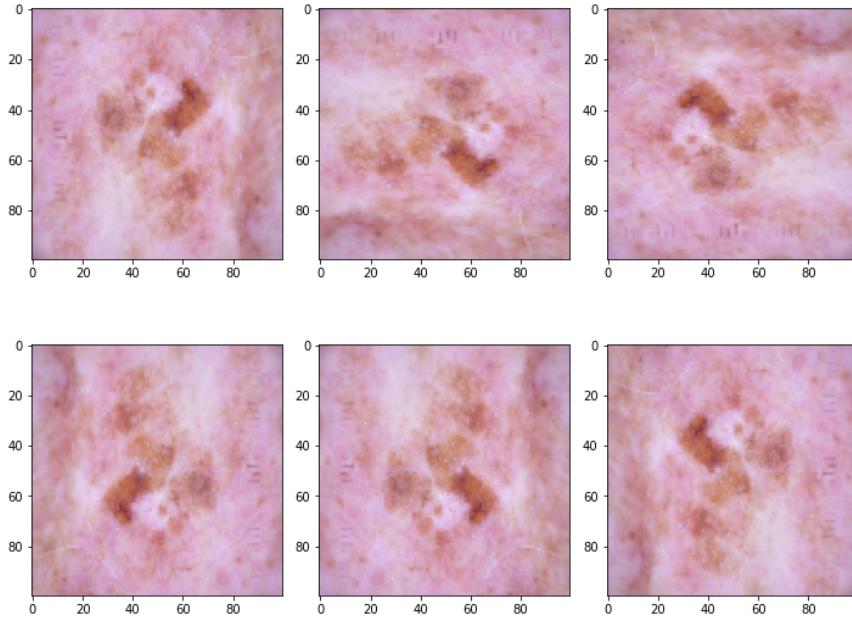


Figure 7: New Images

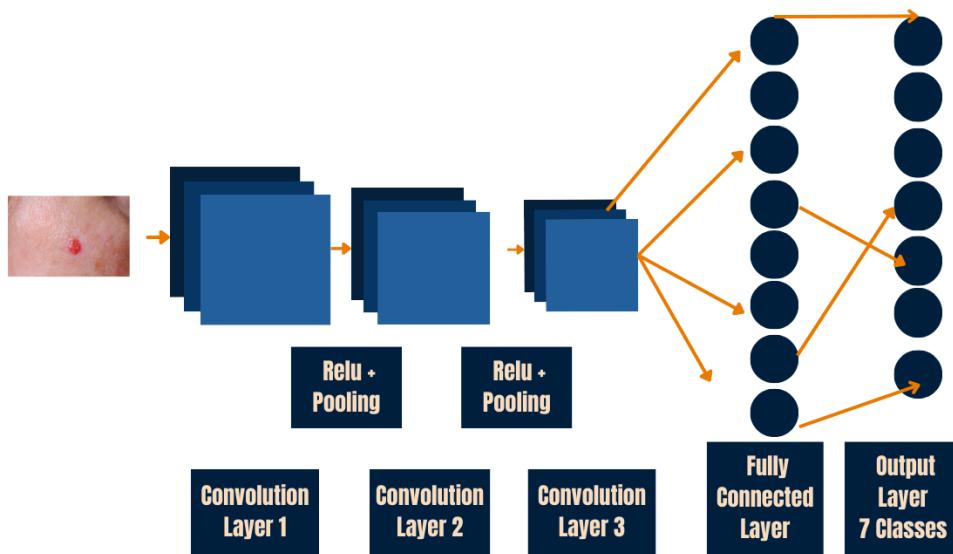
For model definition we define a Convolutional Neural Network model using Keras. Another model architecture is also created. This one includes three convolutional layers on the back of which there is a

corresponding layer of the max-pooling and an additional layer of the dropout for regularization purposes. There is also a final fully-connected layer module that is then followed by diverse softmax activation which made multiclass classification possible on the test and train datasets.(Figure 8 Figure 9)

The dataset so obtained is later separated into the training and the testing data sets. After that, scale the pixel values and set the model with categorical cross-entropy loss and the Adam optimizer for updating the weights. If we believe that the actual implementation of the training loop is missing in this code, then we are working under the wrong assumption.

Lastly, a ResNet50 pre-trained model will be employed for skin cancer classification where predictive analysis will be performed using transfer learning. This data-efficient model needs to be fine-tuned by replacing the fully connected layers of the ResNet50 with new ones that are specifically trained for this task at hand.

Figure 8: Model Diagram for CNN using 300 samples



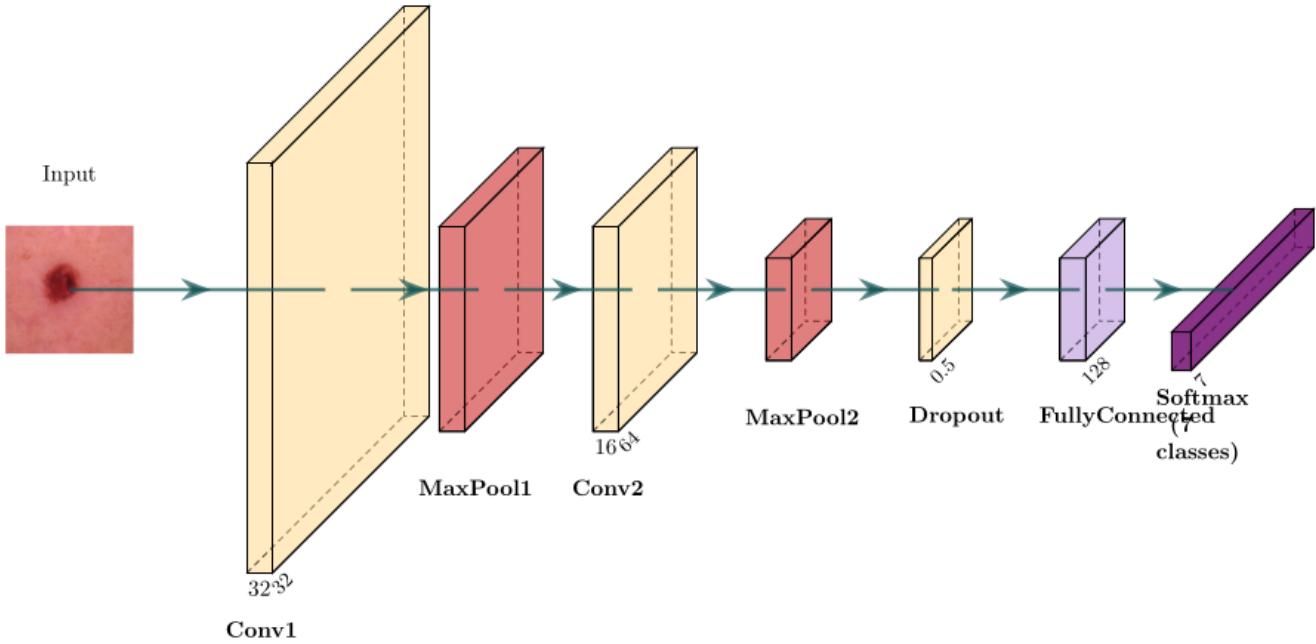


Figure 9: Architecture Diagram For CNN using 300 samples

The model architecture consists of several layers:

Three convolutional functions ordered sequentially. Every stratum is made up from 2D convolutional filters (256, 128, and 64 each) having a 3x3 kernel size. Non-linearity is the concept to capture prominent feature in input images is brought by ReLU activation functions. These blocks are crucial in not only reflecting the spatial patterns but also in defining the atmospheric quality with various attributes. For each convolutional layer, a max-pooling layer is added which is taken apart by a smaller feature image patching and then taking the maximum value from each feature map set. This operation of down-sampling aids in the retention of significant features while preserving the computational complexity. Here, max-pooling with a window size of 2x2 is used. From both regularization and model generalization perspectives, dropout layers are inserted after each pooling layer. A dropout rate of 0.3 is used, whereby 30% of the neurons in these layers are randomly deactivated during training. This is meant to make the network learn robust features. Two fully connected dense layers come after the flatten layer. The first dense layer has 32 neurons that promote learning of higher-level features. The second dense layer has seven neurons, which is the output classes number. With this architecture, a multi-class classification task with seven different classes is implemented.

The training procedure commences by applying the batch size parameter as **16** and setting the number of epochs as **50** at step 1. Considering that this is what this architecture provides, it stimulates the model to be intelligent enough to ensure that samples are processed and covered by sufficient iterations through the dataset. Ensuring that the fit technique that Keras features is set to be training the model using the training data (x_{train} , y_{train}), as well as it will be having a check on the model performance is that will be done continuous via parameters. In the model's training phase the function evaluate (known as the evaluation of model's test accuracy) is used to get important information regarding training success criteria, namely metrics of test accuracy.

Table 3: Base Model Showing Summary

Conv5_block3_2_bn	(none,1,1,512)	2048	Conv5_block3_2_conv[0][0]
Conv5_block3_2_relu	(none,1,1,512)	0	Conv5_block3_2_bn[0][0]
Conv5_block3_3_conv	(none,1,1,2048)	1050624	Conv5_block3_3_conv[0][0]
Conv5_block3_3_bn	(none,1,1,2048)	8192	Conv5_block3_3_bn[0][0]
Conv5_block3_a_dd	(none,1,1,2048)	0	Conv5_block3_2_add[0][0]
Conv5_block3_out	(none,1,1,2048)	0	Conv5_block3_2_out[0][0]
Total params:23,587,712			
Trainable params:0			
Non-trainable params:23,587,712			

Table 4: Model Architecture

LAYER(TYPE)	OUTPUT SHAPE	PARAM #
Conv2d_4(conv_2D)	(None,30,30,256)	7168
Max_pooling2d_3	(None,15,315,256)	0
Dropout_3(dropout)	(None,15,15,256)	0
Conv2d_5	(None,13,13,128)	295040
Max_pooling2d_4	(None,6,6,128)	0
Dropout_4(dropout)	(None,6,6,128)	0
Conv2d_6	(None,4,4,64)	73792
Max_pooling2d_5	(None,2,2,64)	0
Dropout_5(dropout)	(None,2,2,64)	0
Flatten_1(flatten)	(None,256)	0
Dense_2(dense)	(None,32)	8224
Dense_3(dense)	(None,7)	231
Total params:384,455		
Trainable params:384,455		
Non-trainable params:0		

The second model used in this research is Densenet. DenseNet is an acronym used for denoting deep learning architecture that is known for its dense pattern of connections between its layers. Compared to regular convolutional neural networks, where each layer is linked to those succeeding it, DenseNet has direct connections from any given layer to all following layers, and this connection increases feature reuse and information flow across the whole network.

Here, in a DenseNet architecture, each layer receives feature maps from all the preceding layers as a concatenated input. This kind of dense connectivity pattern reuses features, alleviates the vanishing-gradient problem, and favors the efficiency of parameters with their reduced number compared to traditional CNNs. Core layers are denser, such as compressive layers called block layers, which are used for transition and global pooling layers. Batch normalization and non-linearity activation function are used to create a dense block that

integrates a convolutional layer comprising several layers as part of a feature extraction module. Transition layers consist of convolution and pooling operations where dense maps are controlled with cubic reduction in the dimensions of neighbouring slices.. Global pooling layers aggregate spatial information across feature maps to generate predictions by the model based on the whole input.(see Figure 10)

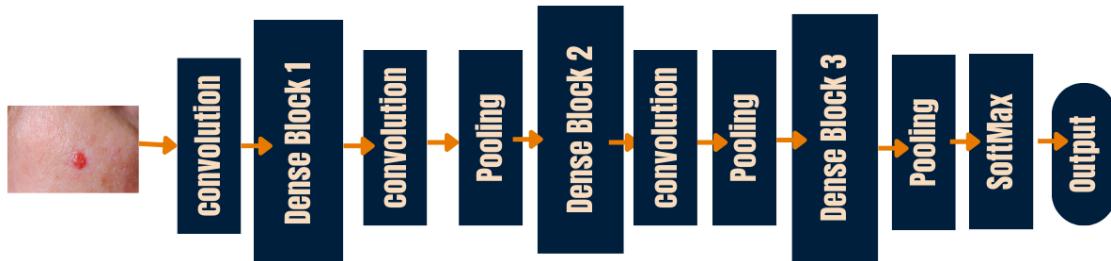


Figure 10: CNN-Lesion_HAM10000

The architectures of DenseNet (Figure 11) show state-of-the-art results in almost every computer vision task like image classification, object detection, and semantic segmentation, with the great efficiency in the usage of parameters and feature reuse. Furthermore, DenseNet has a complex nature which permits it to support modification and extension for special needs, for example, inclusion of attention mechanism and interpretability increase.In this sense, DenseNet indeed is a ground-breaking architecture for CNNs that provide a potent framework for feature extraction, parameter efficiency, and performance optimization.

Training DenseNet neural network model will be performed using the image data generation techniques with the dataset X_train and y_train. Training the model will be done for a particular number of epochs of 100 and with a batch size of 32. During the training of the model, the performance will be enhanced using the help of the Adam optimizer and at the same time a learning rate of 0.0001 is applied, and at this time a categorical cross-entropy loss function will also be utilized, and the metrics for the accuracy will also be considered. An early stopping and model check-pointing callback will be applied to avoid overfitting and to save the best-performing model during training. Finally, the performance of the trained model will be tested using the validation dataset X_test and y_test.

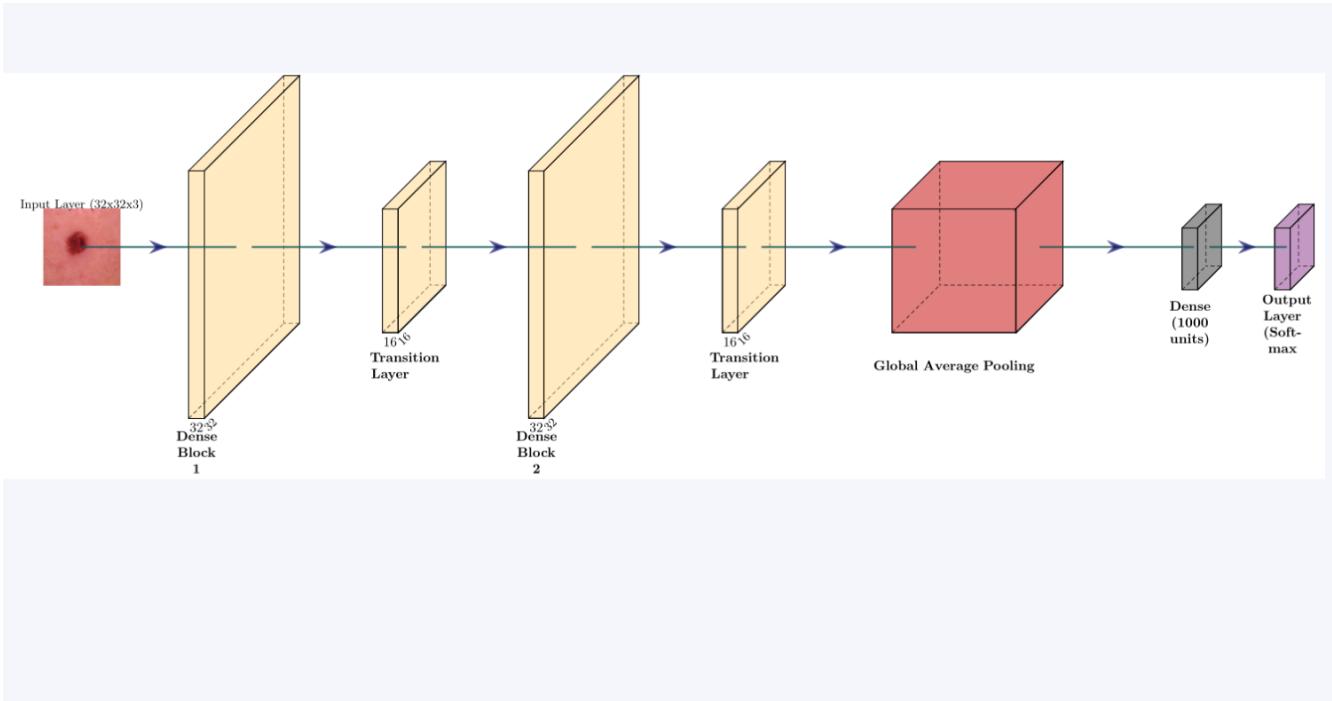


Figure 11: Architecture Diagram of DenseNet

RESULTS AND DISCUSSIONS

The results for CNN model are as follows:

The Figure 12 represents a graph of the learning and validation loss vs. epochs. This graph itself images the output of a regression model that indicates percentage of error being minimized during the training and the generalizing capability of the model on new data, as shown by the validation loss.

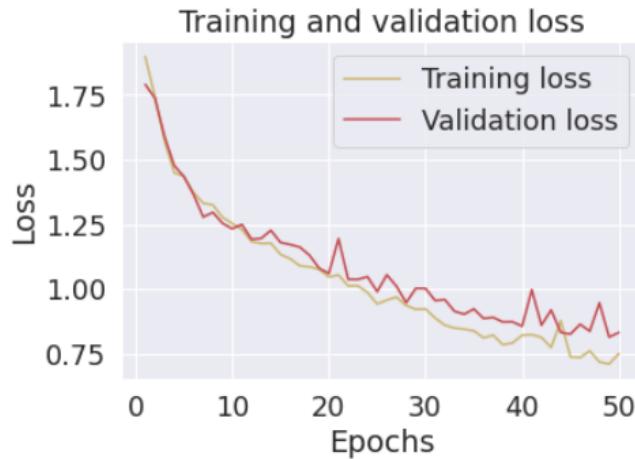


Figure 12: Training and Validation Loss for CNN

Figure 13 shows a line plot of training and validation accuracy over epochs. It provides an insight into how the model is learning in training and how generalizing to new data is captured by the validation accuracy.

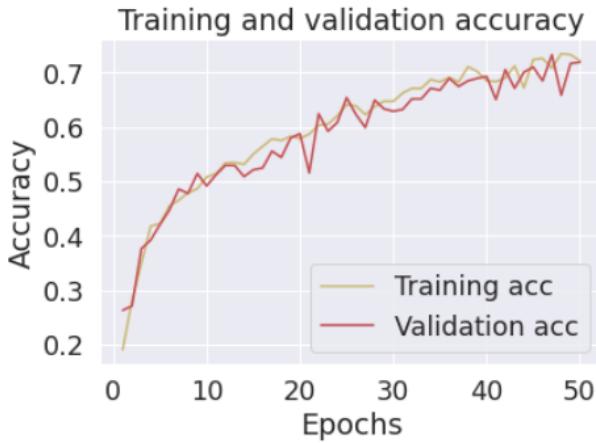


Figure 13: Training and Validation Accuracy for CNN

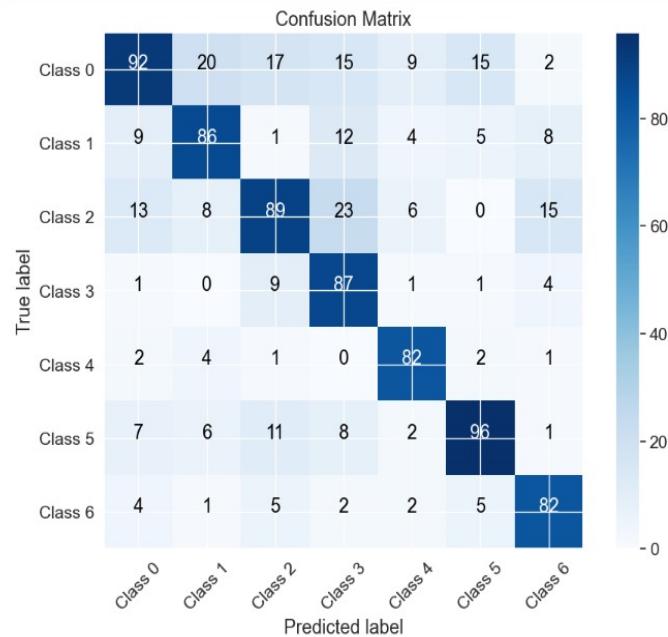


Figure 14: Confusion Matrix of CNN

Figure 14 shows a confusion matrix, which shows the performance of the classifier on the test data. Essentially, the confusion matrix is a breakdown of the model's predictions against the true labels across multiple classes, so it allows us to be able to identify the elements of correct classification and types of error.

Figure 15 shows a plot that shows the fractional incorrect misclassifications for each true label, providing insight into the relative error rates across different classes.

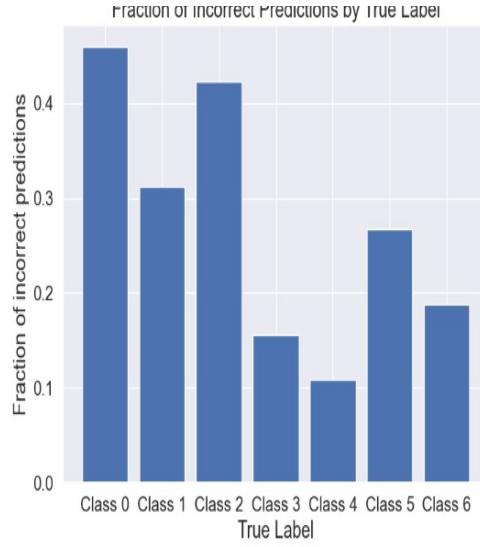


Figure 15: Fractional Incorrect Misclassifications for CNN

TABLE 5: ACCURACY RESULTS OF NEWCNN3500

Epoch	Loss	Accuracy	Validation Loss	Validation Accuracy
1/50	1.8989	0.1905	1.7913	0.2629
2/50	1.7450	0.2811	1.7380	0.2709
3/50	1.5733	0.3486	1.5896	0.3760
4/50	1.4494	0.4175	1.4772	0.3920
47/50	0.7605	0.7082	0.8367	0.7326
48/50	0.7173	0.7345	0.9464	0.6583
49/50	0.7095	0.7326	0.8136	0.7166
50/50	0.7494	0.7215	0.8135	0.7189
Test accuracy: 0.7188571691513062				

```

28/28 [=====] - 1s 35ms/step
      precision    recall  f1-score   support
0       0.80     0.63     0.71     149
1       0.74     0.61     0.67     128
2       0.44     0.70     0.54     120
3       0.93     0.88     0.90     128
4       0.53     0.66     0.59     112
5       0.81     0.55     0.65     113
6       0.98     0.98     0.98     125

accuracy                           0.72      875
macro avg       0.75     0.71     0.72      875
weighted avg    0.75     0.72     0.72      875

```

Figure 16: Classification report for CNN

The results for DenseNet are as follows:

TABLE 6: ACCURACY RESULTS OF DENSENET

Epoch	Loss	Accuracy	Validation Loss	Validation Accuracy
1/100	1.9491	0.3445	2.1044	0.2749
2/100	1.7450	0.4875	1.9979	0.2781
3/100	1.1981	0.5141	1.8699	0.2828
4/100	1.1080	0.5446	1.2588	0.4508
97/100	0.0932	0.9471	0.4920	0.8445
98/100	0.0946	0.9495	0.3162	0.9020
99/100	0.1173	0.9436	3.1451	0.4802
100/100	0.1448	0.9304	0.6086	0.8192
val_accuracy did not improve from 0.90202				

show plots: These plots show the training and validation accuracy of the model (top) and loss (bottom) across the epochs, thereby demonstrating the learning and generalization capabilities of the model.

The Confusion Matrix is a more detailed performance evaluation for the skin lesion classification model: (Figure 20). The confusion matrix given in the image acts as the metric to grade the machine learning model. It consists of four values: true positive, true negative, false positive, and false negative. These values, however, validate the accuracy of the models' predictions. Hence, from the confusion matrix, it can be seen that the model is highly correctly predicting the label "mel." The matrix demonstrates advantages and also places the areas for improvement in order to provide the possibility of additional improvements.

and Figure 20 show a 4x4 grid of subplots with the image from the test dataset, its expected, and predicted labels and the probability of the predicted label.

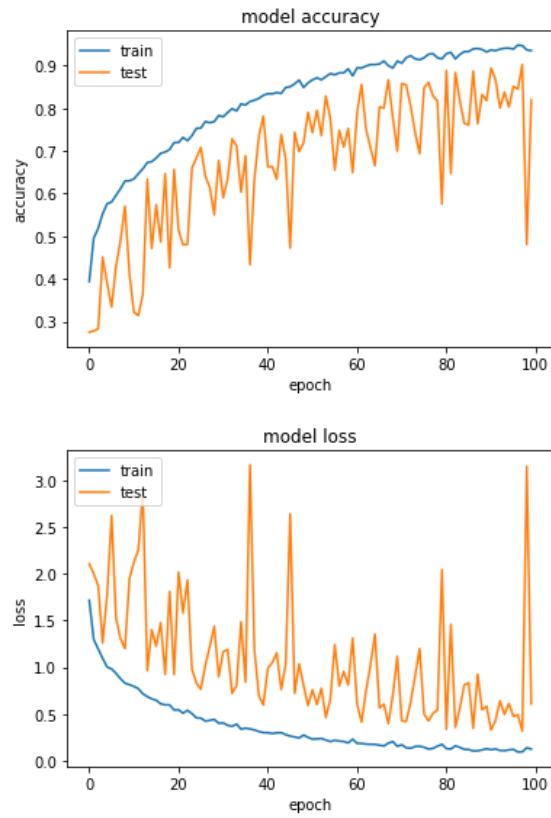


Figure 17: The DenseNet Model's Training and Validation Accuracy (top) and Loss (bottom)

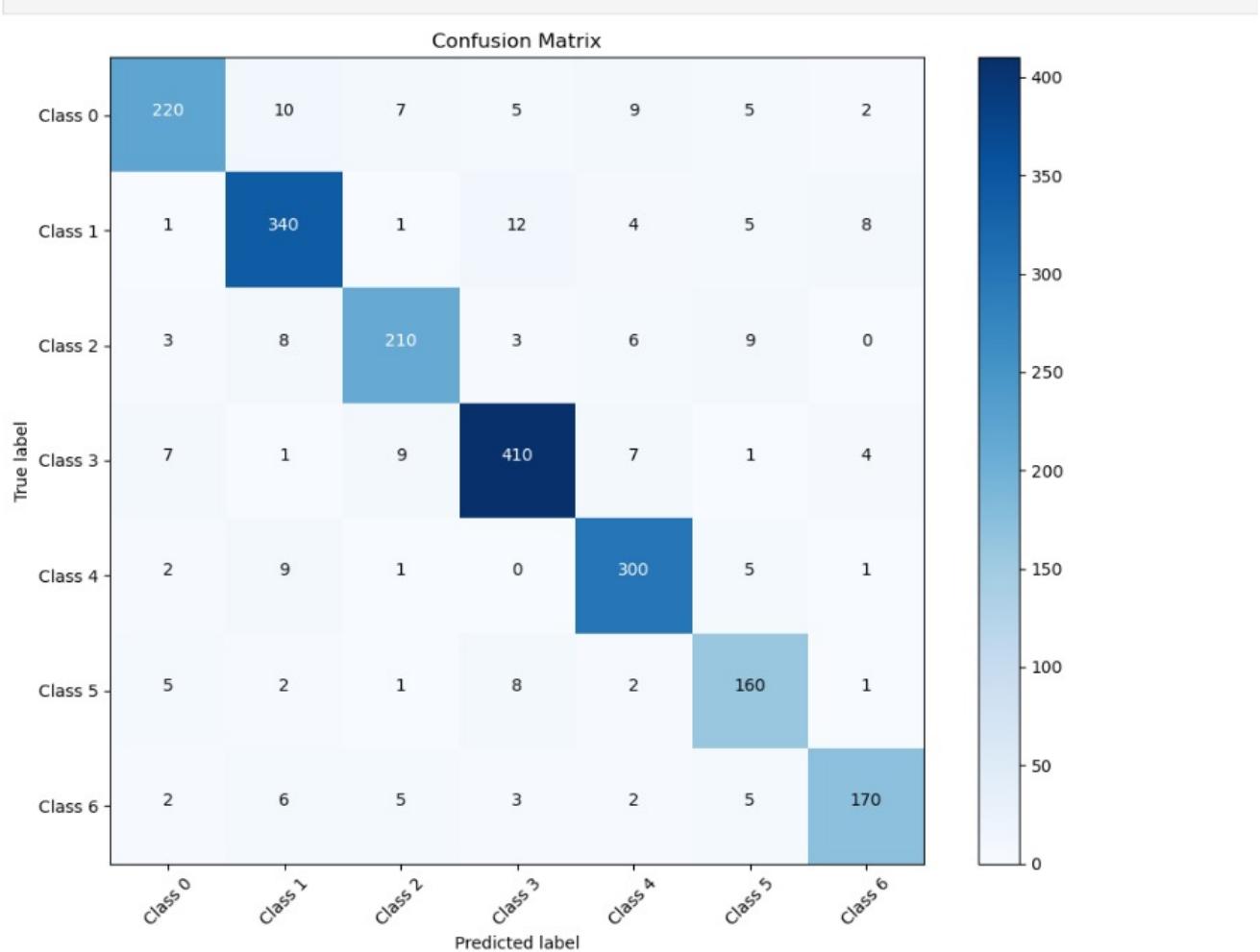


Figure 18: Confusion Matrix

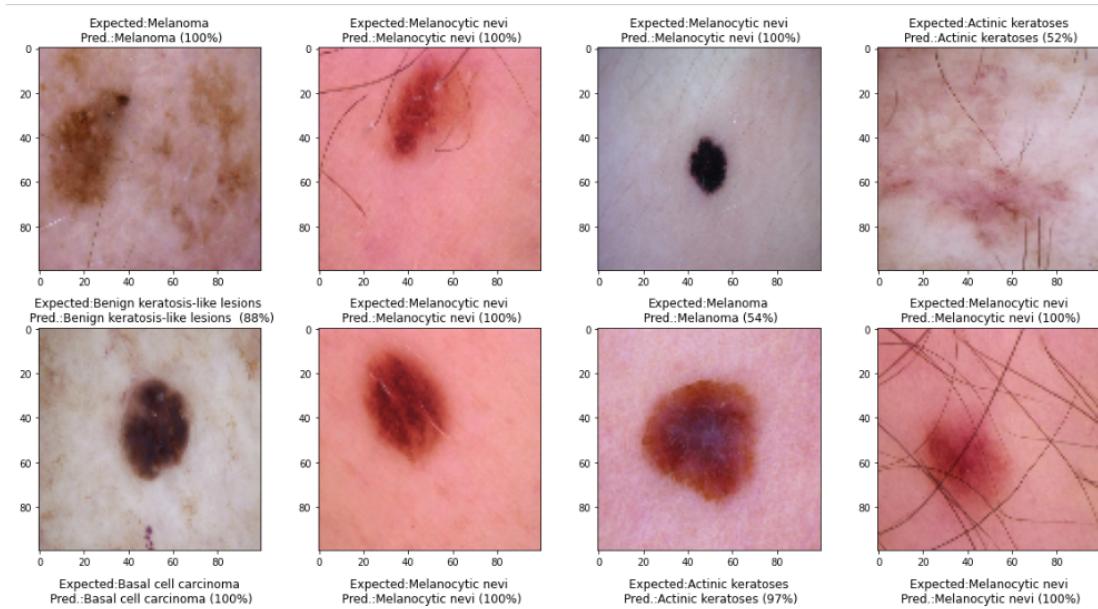


Figure 19: 4 X 4 Grid of Subplots using DenseNet

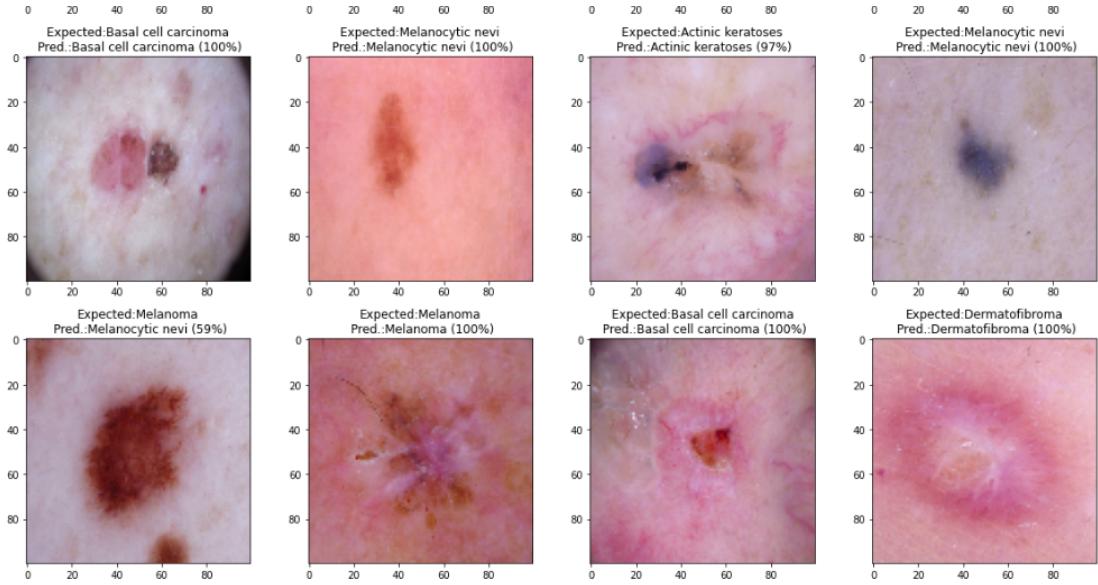


Figure 20: 4 X 4 Grid of Subplots using DenseNet

	precision	recall	f1-score	support
Class 0	0.92	0.85	0.88	258
Class 1	0.90	0.92	0.91	371
Class 2	0.90	0.88	0.89	239
Class 3	0.93	0.93	0.93	439
Class 4	0.91	0.94	0.93	318
Class 5	0.84	0.89	0.87	179
Class 6	0.91	0.88	0.90	193
accuracy			0.91	1997
macro avg	0.90	0.90	0.90	1997
weighted avg	0.91	0.91	0.91	1997

Figure 21: Classification Report of DenseNet

The accuracy computed for Label Equal to 0 in Table 7 accounts for the accuracy of the model's predictions in determining instances where the true label is 0. It calculates the fraction of the true labels, which was correctly predicted, among the total cases where the true label is 0. A higher accuracy score stands for the model reaching a precision level when it recognizes instances of this type well. Next, the Label Different from 0 calculates the accuracy of model's predictions for the instances, when the True label is not the same as 0. It shows binary classification performance by computing the number of correct predictions among all cases where the real outcome is a non-zero value. This is one of the features of the model that evaluates its ability of predicting instances of classes other than the one that is labeled as 0. From the slice shown above we get an idea that the model correctly recognizes classes despite other classes not correctly classified.(Table 8)

The **weighted average accuracy** for the model on all classes is **0.90169**.

TABLE 7: ACCURACY LABELS

Accuracy for label equal to 0	0.85720
Accuracy for label different from 0	0.917149

TABLE 8: THE ACCURACY FOR EACH CLASS INDIVIDUALLY

Accuracy for Melanocytic nevi	0.86
Accuracy for Melanoma	0.87
Accuracy for Benign Keratosis	0.92
Accuracy for Basal cell	0.97
Accuracy for Actinic Keratosis	0.92
Accuracy for Vascular lesions	1.0
Accuracy for Dermatofibroma	0.96

EQUATIONS:

LOSS:

$$\text{MSE} = \frac{1}{N} \sum_i^N (Y_i - \hat{Y}_i)^2$$

RELU:

$$ReLU$$

$$f(x) = \max(0, x)$$

SOTFTMAX:

$$s(x_i) = \frac{e^{x_i}}{\sum_{j=1}^n e^{x_j}}$$

TABLE 9: COMPARISON WITH EXISTING TECHNIQUES

SR NO	References	Problem	Dataset	Technique	Performance	Comparison
1.	Faouzi Adjed1 & Gardez, 2021	Classification of Skin Cancer	ISIC archives	SVM and LBP	76.1%, with sensitivity and specificity rates of 75.6% and 76.7%	Our technique has High classifier learning rates, represent all variations of skin
2.	Kenneth Thomsen1 & Winther2, 2020	Deep Learning for Diagnostic Binary Classification of Multiple-Lesion.	16,453 clinical images from Aarhus University Hospital.	CNN	77%	Our technique Deals with Potential overfitting .
3.	Adria Romero et al. 2017	Classification of skin lesion images	ISBI 2016 Challenge dataset	CNN, VGGNet	79.74%	Our techniques are scalable.
4.	Dan Popescu, Mohamed El khatib and Loretta Ichim, 2022	Classification of skin lesion images	HAM100 00 dataset	CNN	86.71%	Our techniques are optimized.
5.	Rishu Garg, Saumil Maheshwari and Anupam Shukla, 2019	Classification of skin lesion images	MNIST HAM 10000	CNN	90.51%	Our techniques handles the Prediction of result classification
6.	Mohammed A. Al-masni, 2020	convolutional networks for skin lesion segmentation and classification	ISIC datasets from 2016 to 2018	Inception-v3, ResNet-50, Inception ResNet-v2, and DenseNet-201	81.79%	Our techniques Handled imbalanced data .
7.	Javaid et al. 2021	skin cancer classification method using image processing and machine learning	ISIC-ISBI 2016	Quadratic Discriminant, SVM (Medium Gaussian), and Random Forest	93.89%	Our techniques used a different dataset.
8.	Amirreza Mahbod, Gerald Schaefer, Chunliang Wang, Rupert Ecker , Isabella Ellinger, 2019	Classification of skin lesion images	ISIC 2017	CNN architectures — AlexNet, VGG16, and ResNet-18	83.83% for melanoma and 97.55% for seborrheic keratosis classification	Our techniques handled the Overfitting, Data Dependency.

CONCLUSION AND FUTURE WORK

The research focused on the classification of skin cancer using a deep learning approach its accentuation being on Deep Convolutional Neural Networks (DCNN). Using the HAM10000 dermatoscopic image dataset, the study revealed that the proposed DL model improves upon such benchmark models as VGG16, and VGG19, regardless of the evaluation criteria, whether it be accuracy or loss.(Naeem *et al.*, 2022) This advancement is expected to improve diagnoses, as well as the detection of skin cancer diseases.

The better performance of the DCNN model is due to the ability of this approach as well as the inclusion of image augmentation methods.(Pham *et al.*, 2018) The DCNN was developed from the architecture with no competitors allowing the precise selection of the required features that are appropriate for the images of dermatological nature.(Tschandl *et al.*, 2019) Also, the problem of dataset bias, inevitable in medical imaging, was partially solved thanks to the application of image augmentation, including advanced augmentation.(Varoquaux & Cheplygina, 2022) By introducing these modifications the learning process occurred at a faster rate from the existing data, giving better classification accuracy.

For future studies, it will be interesting to apply the DCNN model to various datasets to ensure that method is not dataset-specific. Expanding the patient data to other features could help improve the accuracy of the diagnostic model such as the patient's age and gender or the medical history. Further research based on the combination of various types of machine learning models and other post-processing techniques of X-ray images might result in the creation of more effective diagnostic tools.(Dan *et al.*, 2024) Also, it will be necessary to consider other biases, for example, related to skin type, which could also limit the model's generalizability across populations.

Nevertheless, the presented research has several drawbacks. Specifically, the evaluated model might not be generalizable to other datasets or to actual real-life clinical applications due to the dependency on one dataset known as HAM10000. Moreover, the study is mainly concentrated on the classification of two main categories of skin cancer; melanoma and non-melanoma skin cancer. Additionally, the study didn't consider the issue of racial bias in data samples, which can lead to errors in the model for different races.

Thus, the research itself has the following advantages and limitations: By identifying these limitations, the work proposes potential ways to develop diagnostic technology in the sphere of dermatological disorders.

CITATIONS AND REFERENCES

- 1) Al-Masni, M. A., Kim, D.-H. & Kim, T.-S. (2020). Computer methods and programs in biomedicine 190, 105351.
- 2) Albahar, M. A. (2019). IEEE Access 7, 38306-38313.
- 3) ANSARI, U. B. & SARODE, T. (2017). INT. RES. J. ENG. TECHNOL 4, 2875-2881.
- 4) Garg, R., Maheshwari, S. & Shukla, A. (2019). arXiv preprint arXiv:1912.03798.
- 5) Garg, R., Maheshwari, S. & Shukla, A. (2021). Innovations in Computational Intelligence and Computer Vision: Proceedings of ICICV 2020, pp. 578-586. Springer.
- 6) Javaid, A., Sadiq, M. & Akram, F. (2021). 2021 international Bhurban conference on applied sciences and technologies (IBCAST), pp. 439-444. IEEE.
- 7) Kassem, M. A., Hosny, K. M., Damaševičius, R. & Eltoukhy, M. M. (2021). Diagnostics 11, 1390.
- 8) Lopez, A. R., Giro-i-Nieto, X., Burdick, J. & Marques, O. (2017). 2017 13th IASTED international conference on biomedical engineering (BioMed), pp. 49-54. IEEE.

- 9) Mahbod, A., Schaefer, G., Wang, C., Ecker, R. & Ellinge, I. (2019). ICASSP 2019-2019 IEEE Signal Processing (ICASSP), pp. 1229-1233. IEEE.
- Pavithra, A. & Geetha, B. (2023). AIP Conference Proceedings. AIP Publishing.
- 10) Popescu, D., El-Khatib, M. & Ichim, L. (2022). Sensors 22, 4399.
- 11) Ramachandro, M., Daniya, T. & Saritha, B. (2021). 2021 Innovations in Power and Advanced Computing Technologies (i-PACT), pp. 1-7. IEEE.
- 12) Said, R. A., Raza, H., Munneer, S., Amjad, K., Mohammed, A. S., A. S., Akbar, S. S., Zonain, M. & a. Aslam, M. A. (2022). 2022 International Conference on Cyber Resilience (ICCR), pp. 1-11. IEEE.
- 13) Shetty, B., Fernandes, R., Rodrigues, A. P., Chengoden, R., Bhattacharya, S. & Lakshmanna, K. (2022). Scientific Reports 12, 18134.
- 14) Thomsen, K., Christensen, A. L., Iversen, L., Lomholt, H. B. & Winther, O. (2020). Frontiers in medicine 7, 574329.
- 15) Thomsen, K., Christensen, A. L., Iversen, L., Lomholt, H. B. & Winther, O. (2020). Frontiers in medicine 7, 574329.
- 16) Shetty, B., Fernandes, R., Rodrigues, A. P., Chengoden, R., Bhattacharya, S. & Lakshmanna, K. (2022). Scientific Reports 12, 18134.
- 17) Dan, Q., Xu, Z., Burrows, H., Bissram, J., Stringer, J. S. & Li, Y. (2024). *NPJ Precision Oncology* **8**, 21.
- Naeem, A., Anees, T., Fiza, M., Naqvi, R. A. & Lee, S.-W. (2022). Sensors **22**, 5652.
- 20) Pham, T.-C., Luong, C.-M., Visani, M. & Hoang, V.-D. (2018). *Intelligent Information and Database Systems: 10th Asian Conference, ACIIDS 2018, Dong Hoi City, Vietnam, March 19-21, 2018*,
- 21) Tschandl, P., Rosendahl, C., Akay, B. N., Argenziano, G., Blum, A., Braun, R. P., Cabo, H., Gourhant, J.-Y., Kreusch, J. & Lallas, A. (2019). *JAMA dermatology* **155**, 58-65.
- 22) Varoquaux, G. & Cheplygina, V. (2022). *NPJ digital medicine* **5**, 48.

,