Pakistan Journal of Engineering and Technology
ISSN (Print): 2664-2042; ISSN (Online): 2664-2050
Received: 2024-10-23
Accepted: 2024-12-10
https://doi.org/10.51846/vol7iss4pp183-191

PAKJET
Pakistan Journal of Engineering & Technology

Article

Artificial Intelligence-Based Skin Lesion Analysis and Skin Cancer Detection

Momina Qureshi *, Muhammad Athar Javed Sethi, and Sayed Shahid Hussain

Department of Computer System Engineering, UET Peshawar, Peshawar, Pakistan
* Correspondence: Momina Qureshi (mominaqureshi209@gmail.com)

Abstract: Advanced diagnostic methods are necessary for the early and precise diagnosis of skin cancer, a deadly disease that poses a danger. The accuracy of manual skin lesion assessment and visual inspection is limited, which is why sophisticated diagnostic tools are required. In response, this study presents a groundbreaking approach that makes use of an ensemble of twelve pre-trained deep learning models, including InceptionV3, VGG16, VGG19, Xception, DensNet121, DensNet201, ResNet152V2, MobileNet, MobileNetV2, ConvNeXtLarge, NASNetMobile, and InceptionResNetV2. This study demonstrates a distinct training strategy by employing a two-phase approach: first, training only the newly added dense layers while maintaining the layers of the base model frozen, and then, fine-tuning the entire model. This sophisticated process improves CNN convolutions' stability during feature extraction, which in turn improves the model's overall performance in terms of prediction accuracy. The HAM10000 dataset was used as the main basis for training, evaluating, and comparing all of the models used in this comprehensive research, assuring a consistent and exacting method to progress the field of skin cancer classification. The model with the highest classification accuracy, ResNet152V2, with an F1 score of 98%, wins. By recognizing the intricacy of skin lesions, the study makes the significance of its findings clear and provides hope for the development of more advanced diagnostic instruments. This article not only offers a critical assessment of current methods but also tackles problems and indicates future directions for future research in the field of medical image categorization. This research has implications that extend beyond skin cancer diagnosis; it impacts several therapeutic applications and provides a solid foundation for further advancements in the field.

Keywords: ConvNeXtLarge, DensNet121, DensNet201, HAM10000 Dataset, Inception V, InceptionResNetV2, MobileNet, MobileNetV2, NASNetMobile, ResNet152V2, Skin Cancer, VGG16, VGG19, Xception

1. Introduction

In the rapidly developing science of computer vision, image categorization [1] and detection [2] are at the forefront, creating new opportunities and completely changing the way we handle visual data [3]. These domains have reached previously unheard-of heights thanks to the unrelenting advancement of technology, which has made it possible for robots to accurately perceive, evaluate, and interpret the visual environment [4]. The skin is the body's exterior layer and the body's biggest organ. The skin's ectodermal tissues, which may have up to seven layers, protect the internal organs, muscles, bones, and ligaments underneath. Skin permits the senses of touch, cold, and heat in addition to acting as a barrier between the human body and the external environment. Skin lesions are characterized as patches of skin that are aberrant in relation to normal skin. The first and most basic cause of skin lesions is skin infection. Skin lesions are separated into two categories [5]: the first category, comes with birth or develops over time, and the second category, is brought on by improper handling of the primary skin lesions and can result in skin cancer. Every year, more than three million individuals in the US are confirmed to have skin cancer [5]. During the last ten years, skin cancer has increased in frequency to rank among the most prevalent cancers. [6]. The skin is the biggest organ in the human body, so it seems reasonable that skin cancer would be the most common kind of cancer overall [7]. Malignant cancer and non-malignant cancers of the skin are the two most common classifications [8]. Melanoma is a kind of skin cancer that is severe, uncommon, and occasionally fatal. Despite its higher fatality rate, Melanoma only makes up 1% of all cases of skin cancer, according to statistics from the American Cancer Society [9]. Melanoma develops in cells called melanocytes. Malignant Melanoma develops when normally reproducing melanocytes multiply out of control. It is indeed possible that this will influence the overall body. All the areas exposed to direct sunlight are common sites of development. If not diagnosed early, Melanoma and other skin cancers may spread to other parts of the body, killing the patient painfully over time [10]. Lentigo maligna, acral lentiginous, and nodular Melanoma are subtypes of Melanoma [8]. Non-melanoma categories including basal cell carcinoma (BCC), squamous cell carcinoma (SCC), and sebaceous gland carcinoma, encompass most cancer cases. In the middle and upper epidermal layers, BCC, SGC, and SCC, respectively, are generated. These cancer cells from one part of the body are unlikely to spread to another. Non-melanoma cancers are often less difficult to treat than melanomas. The National Cancer

Institute claims that among all the dangerous malignancies in the world, the most prevalent type of cancer, skin cancer, is diagnosed with more instances annually in the United States than all other cancer types combined. [11]. According to statistics, problems from skin cancer result in the deaths of more than two people every hour. Predictive data estimate an increase in melanoma cases with new diagnoses of 5.8% and a decrease in mortality from Melanoma of 4.8% in 2021 [12]. Invasive melanoma cases have increased by 44% over the past ten years, and it is observed that more than 7000 people died from the disease in 2021 [8]. More than 5400 fatalities per month are through in-time detection is thus crucial for skin cancer treatment [13]. Physicians to find cancer often apply the biopsy approach. With this treatment, a physician separates a part of a potentially malignant skin lesion for examination. This is a tedious and time-consuming process. The manual procedures take a lot of time and depend on the operator, which might lead to mistakes or incorrect illness diagnoses. The study of skin issues is greatly aided by an automatic diagnosis of skin cancer, which significantly lowers the expense and effect of human detection. It aids in the early diagnosis of skin malignancies. A manual skin cancer diagnosis is a tedious, time-consuming process, and costly. As AI technologies are becoming faster and smarter, it is not surprising that they are being used to help identify skin cancer and prescribe therapies [14]. This is due to the widespread belief that AI-based methods are inexpensive, easy to use, and readily accessible. Compared to other diseases, skin cancer diseases have a much higher fatality rate. Yet, an early diagnosis could help. Yet, many skin malignancies first show no signs. Medical specialists highly advise that the examination be carried out at a certain age in regions where skin malignancies are particularly prevalent. The time and expertise a doctor requires to diagnose skin cancer precisely and effectively during a physical examination are highly sensitive, and the growing number of examination screenings causes the doctor to get fatigued and more prone to making mistakes. Having an automated AI-based model that serves as a decision support system for healthcare is very helpful. Consequently, the suggested issue in this research study is the necessity for an intelligent computer-aided detection system to identify skin cancer early.In recent years, deep learning has significantly changed the field of machine learning. It is regarded as the most advanced subset of artificial intelligence. The structure and operation of the human brain have an impact on these algorithms. Several different industries employ deep learning techniques, such as bioscience [15], reinforcement learning [16], and voice recognition [17]. Comparing deep learning systems to other conventional machine learning techniques in various domains has yielded good results. Several deep learning techniques have surfaced recently and are being used for computer-based skin cancer identification, including generative adversarial neural networks, convolutional neural networks, artificial Neural networks, and Kohonen self-organizing neural networks (KNN) (GAN). In this study, we provide an AI-based method for detecting skin cancer. This study aims to explore and analyze a dataset of human skin images to investigate skin cancer. A predominant focus was placed on the development of an advanced AI-based model specifically designed for accurate diagnosis and classification of different kinds of skin malignancies was given priority. Acknowledging the significance of robustness in the model, great care was taken to fine-tune different parameters.

2. Literature Review

Skin image analysis is still a developing area since skin cancer is such a difficult disease to diagnose. Several researchers through the introduction of new methods are always improving the performance of image classification. The area is now more intriguing and diversified because of technological advancements and the advent of new technologies. Esteva et al. [18] made the first significant advancement in the categorization of skin cancer using a pre-trained Google Inception V3 CNN model. In this research, about 129,450 clinical skin cancer images, including 3,374 dermatoscopic images, were used. The resultant accuracy of classification is 72.1%. On the ISBI 2016 challenge dataset, Yu et al. [19] CNN with over 50 layers was constructed in 2016 for the classification of malignant melanoma cancer. 85.5% was the highest recorded classification accuracy for this study It is essential to provide justification for undertaking the proposed research, perhaps in the light of previous work done. It should be possible in most cases to anticipate the specific and general benefits likely to be achieved as a result of the completion of the proposed research. In 2018, Haenssle et al. [20] reported 86.6% sensitivity and specificity for the categorization of dermatoscopy melanocytic pictures into a binary diagnostic category using a deep convolutional neural network. Dorj et al. used deep learning CNN and ECOC SVM to develop a multiclass classification in a study [21]. AlexNet Deep Learning CNN that had been previously trained, and ECOC SVM were used in the classification of multiclass data. The average accuracy is stated to be 95.1%.

In [22], the authors presented a technique to classify skin cancers into benign and malignant subtypes. The intended system was divided into three phases. In the first stage, lesions were extracted from images using a NN that generates its own data. Details about the tumor's edges, appearance, and colors were fetched in the 2nd stage. Then, an NN ensemble method was utilized to categorize cancers. The accuracy of categorization is enhanced by an ensemble of NN. The outcomes of the suggested classifier were evaluated against the results of various classifiers, including support vector machines, k-nearest neighbors, random forests, etc. With a 91.11% accuracy, the recommended model surpassed the other classifiers by at least 7.5% in terms of sensitivity. In [23], the authors investigated an artificial neural network (ANN). By incrementally modifying its connection eights, the ANN outperformed the

KNN in terms of accuracy by cutting down on the error between the actual outputs and predicted ones. On the other hand, melanoma detection systems are increasingly using deep learning. While typical pattern recognition algorithms primarily rely on the results of the segmentation stage, deep

learning uses a vast number of skin photos to automatically recognize the skin lesion and generate a feature map. A method for automatically detecting melanoma lesions on skin imaging has been suggested in article [24] and is based on the idea of deep learning. The findings demonstrate that when utilizing CNN with a 15×15 training input size, deep learning is effectively able to identify the melanoma lesion. Yet, any kind of skin melanoma may be reliably detected using this neural network design.

A hybrid approach for melanoma skin cancer detection has been presented in [25] that can be utilized to evaluate any unsure lesion. Their recommended solution relies on a majority vote to combine the predictions of three separate methodologies. The only two rules most systems rely on are the ABCD rule and the Blue-Black rule, both of which have been shown to have certain limitations and sometimes to be ineffective. The ugly duckling was another suggestion made. By examining the anatomy of the problematic lesion and contrasting it with those of surrounding lesions, the aim was to identify an outlier among a background of moles that had similar characteristics. Clinical research has shown that this clue is a valid factor for cancer detection, but it has not yet been investigated in the proposed study's autonomous melanoma detection systems. Using hybrid feature extraction, skin cancer has been classified as benign or malignant in this study [26]. Several criteria, such as the ABCD rule, HOG, etc., are used in machine learning techniques to automatically recognize skin lesions to extract features and subsequently classify. The segmentation of the skin lesion using the GAC approach was suggested. For feature extraction, the ABCD rule, appearance of the skin lesion GLCM, form, and boundary of the skin cancer HOG were recommended for color, symmetry, and diameter of the skin lesion.

Several machines learning methods, including SVM, KNN, and Nave Bayes, were proposed to handle the classification. Skin lesion imaging from the ISIC dataset was processed using the recommended method. When all methods of classification are compared, SVM performs better than the others. KNN produced results with 85% specificity and 86.2% sensitivity. This method can also be applied to the neural network platform for increased accuracy. ANN based automated skin lesion diagnosis was suggested [27]. To extract features, this approach used a wavelet transform. Cancerous and noncancerous pictures were identified using the suggested ANN model. Besides, ANN- based skin lesion diagnosis method was presented by Choudhari [28]. Each image's components were identified using maximum entropy thresholding. A gray-level co- occurrence matrix was used to extract the characteristics of the skin lesions (GLCM). Finally, a feed-forward ANN distinguished between skin cancer's malignant and benign stages by analyzing the input images. In contrast to other works of literature that have already been published, the goal of this study is to enhance the early diagnosis system of skin cancer by employing readily accessible data and applying cutting-edge AI computation and algorithms.

3. Materials and Methodology

This section explores the thorough resources and techniques that support our study. Fig. 1 visually illustrates the methodology used for our approach to multiclass classification of skin cancer, which involves a series of essential steps. We start the procedure by obtaining a publically available dataset and then organizing it into a suitable format for further processing. We next carry out the required preprocessing processes to improve the quality and consistency of the sample photos after this data collection and organization phase. In order to determine which AI model performs the best among the competitors, we next train a variety of AI models using preprocessed images as input. The subsequent sections of this section provide a detailed analysis of each of these important phases, shedding light on the precise
actions and strategies employed to meet our goals.

Figure 1. Proposed methodology.

3.1. Dataset

Using the HAM10000 dataset [29], a large collection of dermatoscopic images of pigmented lesions, we investigate the intricate field of dermatology in this thorough study. This dataset provides a solid basis for analyzing the unique properties of different skin lesions, enabling a more accurate diagnosis that surpasses generalizations and provides an exhaustive exploration into the myriad skin problems a patient may encounter.It is difficult to train neural networks for the automatic identification of pigmented skin lesions due to the minimal number and diversity of dermatoscopic image datasets. The HAM10000 ("Human against Machine with 10000 training images") dataset emerges as a game-changing way to tackle this problem head-on. The authors have cleverly gathered a substantial dataset by carefully collecting dermatoscopic images from various demographics and capturing them using a variety of modalities. This collection, which includes 10,015 dermatoscopic pictures in total, is ideal for supporting academic machine learning projects. Each image is 600x450 pixels in size and is presented in the widely used RGB format, providing a broad canvas for study.The data has been carefully organized to include a wide range of important diagnostic categories that fall within the group of pigmented lesions. These classes span actinic keratoses and intraepithelial carcinoma Bowen's disease (akiec), basal cell carcinoma (bcc), benign keratosis-like lesions (solar lentigines seborrheic keratoses and lichen- planus like keratoses, bkl), dermatofibroma (df), Melanoma (mel), melanocytic nevi (nv), and vascular lesions (angiomas angiokeratomas, pyogenic granulomas, and haemorrhage, vasc).

3.2. Exploratory Data Analysis

Fig. 2 provides an effective visual depiction that highlights the significant imbalance found in the dataset. As we examine the distribution of the sample images across different classes, this startling discrepancy becomes even more obvious. Notably, there is a noticeable bias in the composition of the dataset owing to the class designated as "Melanocytic Nevi" outnumbering the other classes. The fascinating conclusion regarding the ground truth annotations in the dataset is revealed in Fig. 3. A thorough investigation finds that a sizeable majority exceeding the 50% cutoff are rooted in the exacting field of histopathology (histo). The strength of the dataset's basis is attested to by these thoroughly verified examples. The authenticity of the remaining cases, which span a wide spectrum and contribute significantly to the dataset's inherent depth, comes from sources like follow-up examination (follow-up), expert consensus (consensus), and even the accuracy of in-vivo confocal microscopy (confocal).

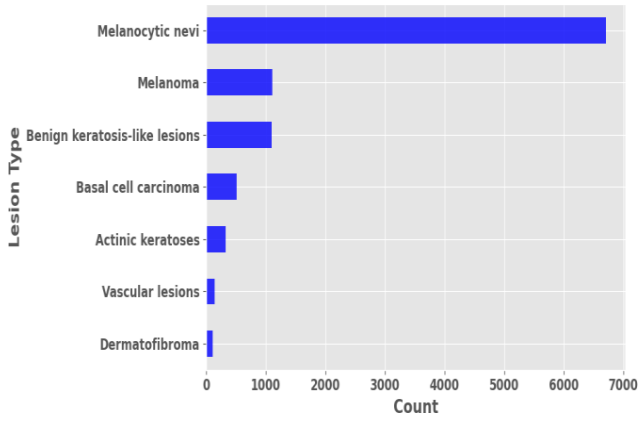

Figure 2. Number of sample images in each class.

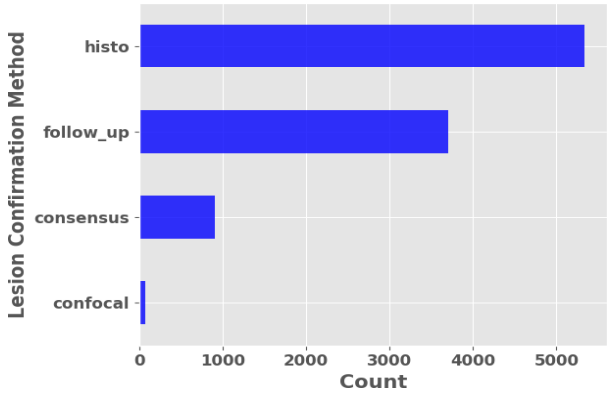

Figure 3. Lesion confirmation methods.

The collection contains a wide variety of lesion images that have been systematically collected from diverse anatomical areas of the human body, creating an extensive mosaic of dermatological samples. Fig. 4 illustrates how this spatial distribution spans a broad spectrum and includes prominent areas including the back, trunk, face, chest, and many more. Due to its geographic diversity, the dataset has a varied fabric that is indicative of the rich variability seen in actual clinical settings. A captivating examination of the complex interaction of age and gender dynamics in the dataset can be seen in Fig. 5 and Fig. 6, respectively. In this thorough research, determining the distribution of sample images among different age groups and genders is a crucial step. By giving a clearer understanding of the interactions between these significant demographic parameters and the occurrence of pigmented skin lesions, this in-depth examination promotes a refined awareness of the patterns of presentation of these lesions.

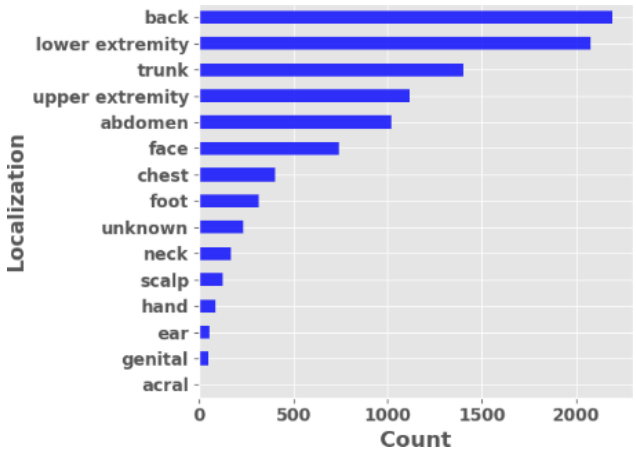

Figure 4. Localization analysis.

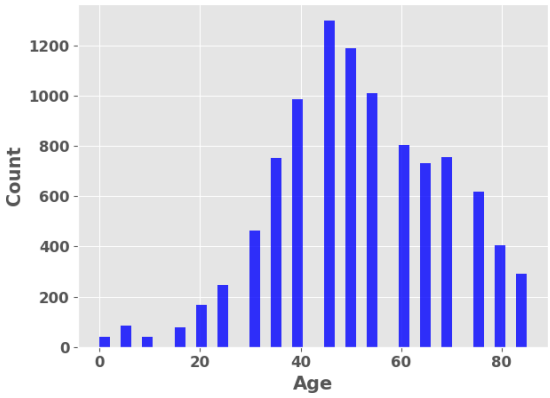

Figure 5: Age analysis.

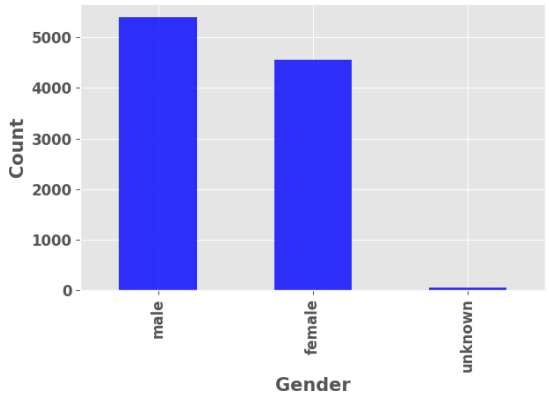

Figure 6. Gender analysis.

This multidimensional study of the dataset's spatial distribution, demographic composition, and localization idiosyncrasies essentially adds a substantial level of complexity. This dataset captures the intricacies of age, gender, and anatomical placements, making it an essential foundation for comprehensive investigations and in-depth study within the complicated subject of pigmented skin lesion analysis.

3.3. Preprocessing

The well gathered dataset used in this paper was obtained from the famous Kaggle platform. The dataset was pre-processed to fit in the structure of our model. The first thing to do was rescale the lesion photos so that they were in dimension of (128, 128, 3). A vital data standardization process that meant scaling integers by 255 to establish a homogeneous planet. Within the TensorFlow deep learning framework, that is acclaimed for its computing strength and advanced mathematical constructions, the complicated step of model training unfolded. Through a random permutation of the dataset, the training set, validation set, and test set effectively materialized. This split was carried out while upholding the rational randomization concept. Notably, the test set was mindfully given 20% of the dataset's size to ensure a strong assessment setting, while the training set, making up an impressive 80%, provided the optimal setting for the development of our neural architecture.

Looking further into the structure of the training dataset, a selective subset—5% to be precise—was chosen for validation purposes, promoting experimentation and optimization of model parameters. A dataset of 12,018 samples was included in the experiments and accuracy, signifying the weighty and comprehensive nature of our study's foundation. In keeping with the pattern of the training data, a calculated partition placed 9,514 samples in the training set's embrace while 501 samples got into the validation set, leading to the creation of a test data reserve with 2,003 skin lesion image samples. This created a delicately fostering setting for the development of our AI architecture.

3.4. AI Models

This study centers on the challenging goal of distinguishing and predicting, with an excellent level of accuracy, the existence of skin cancer e.g., Melanoma—a disease that stands as the deadliest type of skin cancer—using cutting-edge developments in Deep Learning Classification models. In addition to Melanoma, a variety of alternative malignant forms, ranging from melanocytic nevi to a constellation of six different kinds, are also attempted to be identified and classified. This study creates a significant opportunity for comprehensive diagnostic findings. Our goal is to provide medical professionals and the research community with a high accuracy model for the precise detection and categorization of a variety of dermal disorders. This complex purpose highlights our dedication to using the strength of cutting-edge Deep Learning models to promote a transformational influence on dermatological diagnostics, allowing early identification and informed decision-making for better patient outcomes.

In this research, we leverage the power of an ensemble of twelve pre-trained models that have been carefully chosen for their individual strengths and enhance their current architectures to achieve new heights of performance accuracy. VGG16, VGG19, InceptionV3, Xception, DensNet121, DensNet201, ResNet152V2, MobileNet, MobileNetV2, ConvNeXtLarge, NASNetMobile, and InceptionResNetV2 are the models included in this list. We attempt to push the limits of these models through a careful architectural refinement process, ensuring that they serve as examples of cutting-edge technology in this field of image classification. This comparison allowed us to determine that ResNet152V2 was the top-performing model, with an F1-score of 98% and a classification accuracy of 98.95%.

3.5. Training of AI Models

The algorithm's training process was competently divided into two distinct phases, each of which was carefully designed to result in an extensive structure ready for maximum prediction prowess:

3.5.1. Phase 1

In this phase, we focused on enhancing the existing architecture with new dense layers that were strategically set on top of the basic base model. The fundamental principle driving this phase was to strictly retain the layers of the basic model's immutability. A steady output from the CNN convolutions was produced by the strategic constraint in training them that served as a pivot. Simultaneously, the newly added dense layers were used in categorizing the extracted features into their corresponding lesion class.

3.5.2. Phase 2

In this phase, the entire model architecture was encapsulated by a competent injection of fine-tuning. This phase's sole goal was to boost the network's prediction accuracy and bring it to the pinnacle of performance. A reduced learning rate is employed to avoid making drastic modifications to the feature extractors.

The predictions of the entirely untrained dense layers would be nearly random if they were trained from the beginning, i.e., skipping Phase 1 and immediately training the whole CNN. Due to this randomness, there would be a significant loss that would propagate throughout the whole network. The well-trained feature detectors in the convolutional layers, which had previously discovered significant patterns, may become disrupted or broken because of this backpropagation. For this reason, a two-phase technique is utilized to provide steady training while maintaining the caliber of the learned features. The overall model architecture is depicted in Fig. 7.

4. Results

This section gives a thorough description of the methods used to get better results and makes comparisons with comparable

areas of research. The subjects cover a range of AI-based classification models, in-depth analysis for evaluating metrics, and model performance evaluation. Several figures and tables are used to show, explain, and augment the findings of these comprehensive studies, making it easier to grasp the research's conclusions in full.

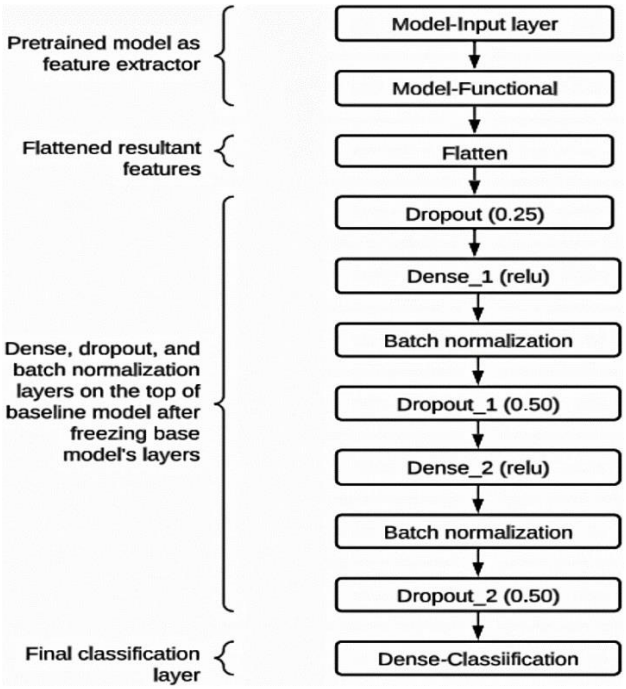

Figure 7. Model Architecture.

4.1. Modelling Parameters

With the primary objective of performing a detailed comparison study, we carefully implemented each of the models in our examination. We enforced these across the complete training pipeline by carefully curating a defined, matching set of hyper-parameters. More specifically, we trained using the Adam optimizer and Sparse Categorical Cross-Entropy Loss to improve model performance. We used a learning rate of 1e-5. Also, we scheduled our training to exploit mini-batch gradient descent with the help of 128-batch We extensively trained both models for 50 epochs each so that they would have plenty of time to converge and learn because the course encourages proper experimentation. We verified that we were able to control how well our model performs while training since every time a full epoch is completed, the performance of the resulting architecture must be assessed on the validation set. These are resource-intensive experiments, and Kaggle gave us the computational power to process more sophisticated models by running them within a managed environment. Not only that, the IDE of Kaggle also comes with a very huge RAM size at 30 GB and we have free access to GPUs such NVIDIA T4. We explain the result of each of these models below.

4.2. Evaluation Metrics

With the test and also on original datasets, which have very less balanced classes, we took a holistic approach to understand our model's performance. We also used the weighted average F1-score as evaluation metric, besides accuracy and precision. We incorporated confusion matrix, Receiver Operating Characteristic Curve (ROC), and Area under the Curve (AUC) into our evaluation methods to ensure a deep review of each model performance. Let's explain these evaluation metrics:

4.2.1. F1 Score

We measured our model in terms of its F1- Score, which is the balance between precision and recall. Recall differs to precision in that the former calculates how many positive cases there are within a dataset, while precision is estimating what proportion of total sayings really is about high-value predictions from algebra-return values. By merging these two together and converting them into a single number, F1-score gives an unbiased measure of how good our model is at correctly classifying positive as well negative examples. The F1 score is the harmonic mean of precision and recall. Mathematically, it is represented as (1)

$$F1\ Score = \frac{\text{Precision} * \text{Recall}}{\text{percision} + \text{Recall}} \quad (1)$$

4.2.2. Accuracy

The f1-score is a measure of the accuracy of the model after precision and recall have been balanced. Precision measures the ratio of actual positive predictions to all other predicted positives generated by the model (in contrast to recall, which tracks how many real true instances are in the data) Averaging these two metrics gives us the value of F1-Score – a balanced assessment on how well does your model perform in classifying positive samples and negative samples together.

4.2.3. F1 Score

F1 Score is the harmonic mean of precision and recall Mathematically represented in (2)

$$Accuracy = \frac{TN + TP}{TN + FP + TP + FN} \quad (2)$$

where true positives are represented by TP, true negatives by TN, false positives by FP, and false negatives by FN.

4.2.4. Precision

When comparing all the model's positive predictions, precision is the percentage of true positive predictions. It evaluates the model's capacity to keep clear of false positive errors. In other words, precision assesses how well the model identifies Positive data. Equation (3) provides precision:

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

4.2.5. Confusion Matrix

A confusion matrix is a table that offers a detailed explanation of the predictions made by a model. Its four values are False Positives (FP), False Negatives (FN), True Positives (TP), and True Negatives (TN). These figures contribute to a more comprehensive comprehension of the model's functionality.

4.2.6. ROC Curve

The ROC curve graphically illustrates a model's capacity to distinguish between the positive and negative classes at various decision thresholds. It displays the true positive rate (sensitivity) against the false positive rate (specificity) as the decision threshold varies. The ROC curve provides insight into how well a model can differentiate between different thresholds.

4.2.7. AUC

The ROC curve yields a numerical number known as the AUC. It measures how well a categorization model performed overall. A perfect classifier has an AUC of 1.0, while a random classifier has an AUC of 0.5. Better model performance in differentiating between positive and negative examples is shown by higher AUC values.

4.3. Performance Evaluation

As part of our analysis, we looked closely at the performance of the model in testing. We considered other critical parameters, namely weighted average F1 scores, accuracy, and precision, all of which are found in their classification reports in Table 1. A key point to mention is that machine learning models should not be based solely on accuracy. The weighted average F1-score gives a fairly complex picture of overall effectiveness, balancing recall and precision. For one, these results abundantly depict the extraordinary performance of ResNet152V2 in this skin cancer classification challenge since it reached an impressive 98% classification accuracy and F1 score, thus shedding light on its superior capabilities in comparison with all the models considered. This high accuracy really does credit to its capability to accurately classify skin cancer images in a variety of classes showing a very good architecture and efficient training of the model.

Table 1: Performance evaluation of the developed models.

Model	Accuracy	Precision	F1 score
VGG16	83.72%	82.76%	82.58%
InceptionV3	79.38%	77.51%	77.32%
Xception	87.42%	87.15%	86.49%
DensNet201	92.36%	92.33%	92.08%
MobileNet	91.01%	91.1%	90.55%
MobileNetV2	86.82%	86.69%	85.63%
ConvNeXtLarge	74.94%	73.93%	69.97%
NASNetMobile	76.39%	74.16%	72.43%
InceptionResNetV2	82.33%	81.48%	80.95%
ResNet152V2	98.95%	98%	98%

4.4. Confusion Matrix

All the confusion matrices may be inspected to gather significant information. Fig. 8 (A-L) is the confusion matrix for all the models. Interestingly, Fig. 8 (L) is the confusion matrix for ResNet152V2, which Stand out as the most enlightening and illuminating model evaluated the given confusion matrix corresponds to the following seven different class labels: actinic keratoses, basal cell carcinoma, benign keratosis-like lesions, dermatofibroma, melanocytic nevi, Melanoma, and vascular lesions.

Correlates with in terms of classification. While looking at the confusion matrix of ResNet152V2 in Fig. 8 (L), the diagonal values in the matrix are true positives (accurate classifications), demonstrating the model's precision for each class. Particularly noteworthy is how well ResNet152V2 performs in accurately classifying Dermatofibroma and Melanocytic nevi with no erroneous predictions. Nevertheless, there are very few instances of misclassification, such as false positives (for example, classifying Melanocytic nevi as Actinic keratoses) and false negatives (for example, classifying Actinic keratoses as Benign keratosis-like lesions). This confusion matrix shows the amount of accurate and wrong predictions for each category and offers a thorough analysis of how ResNet152V2 outperformed each class.

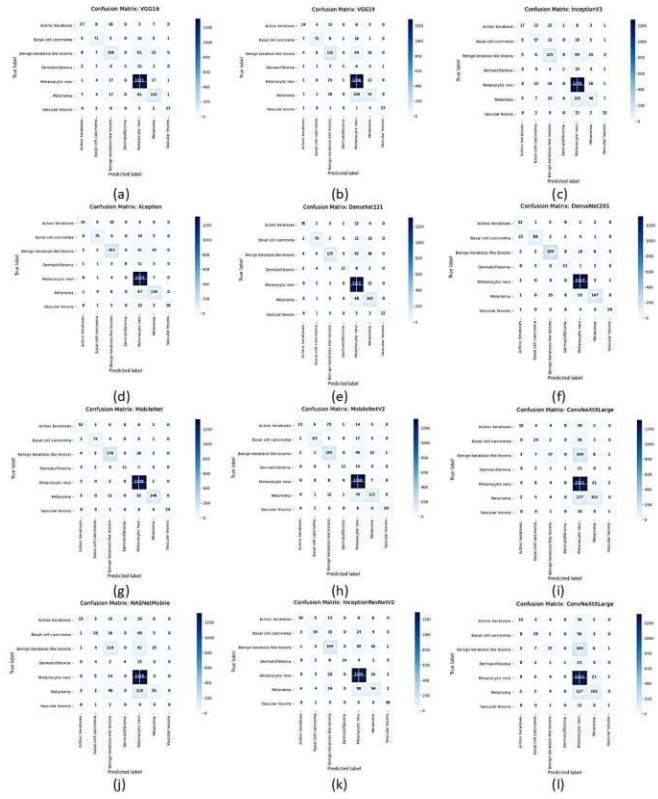

Figure 8: Confusion matrix of the developed model.

4.5. ROC & AUC

In addition to the previously indicated evaluation standards, a look at the ROC curves for each model, as shown in Fig. 9 (A)

through 9(L), provides important insights into their discriminatory skills. Fig. 9 (L) prominently displays the ROC curve for ResNet152V2, which stands out as the strongest model among those examined. Additionally, the AUC values are simply provided by the ROC figures, underlining ResNet152V2's outstanding performance and high AUC score. Also, A comparison of accuracy with existing studies leveraging the HAM10000 dataset is presented in Table 2.

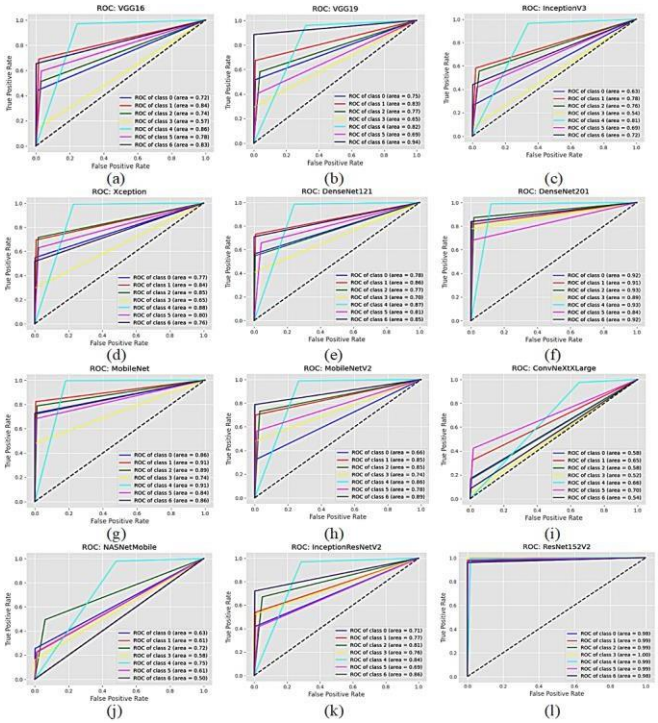

Figure 9. ROC of the developed model.

Table 2: Comparison of accuracy with existing studies leveraging the HAM10000 dataset.

Reference	Year	Algorithm	Accuracy
Waweruet al.[26]	2020	DCNN	78.0%
Huo [27]	2021	CNN	75.0%
Moldovan [28]	2019	Transfer learning	85.0%
Yildiz Aydin [29]	2023	XGBoost classifier	96%
Shah [30]	2021	LRNet	90.6%
Amin Tajerian [31]	2023	EfficientNet	84.3%
This study	2024	ResNet152V2	98.0%

5. Conclusion

This study significantly advances the field of skin cancer categorization by utilizing state-of-the-art Deep Learning Classification models to tackle the difficult task of precisely identifying and predicting seven different types of skin cancer, which are among the deadliest types of the disease. We improved their performance by carefully optimizing the architecture of a specially selected set of twelve pre-trained models, which are: VGG16, VGG19, InceptionV3, Xception, DensNet121, DensNet201, ResNet152V2, MobileNet, MobileNetV2, ConvNeXtLarge, NASNetMobile, and InceptionResNetV2. The HAM10000 dataset provided the fundamental framework for training, evaluation, and comparison of all models used in this extensive study, guaranteeing a consistent and exacting method to progress the area of skin cancer classification. ResNet152V2's outstanding result, featuring an F1 score and an astounding 98% classification accuracy, highlights how effective it is in correctly classifying images of skin cancer. Nonetheless, it is important to recognize certain limitations that are intrinsic to this research. Future research including bigger and more diverse datasets that cover real-world differences will be necessary since the model's generalizability to broader clinical settings may be limited due to its dependence on carefully selected datasets. Furthermore, the complexity of clinical decision-making could not be well captured by an exclusive focus on image-based categorization. Future studies may be carried out to give priority to the integration of multi-modal information, such as clinical data and patient histories to improve the model's diagnostic skills in various scenarios. Iterative model refining is also necessary due to the continual developments in deep learning approaches, which need constant adaption. In addition to strengthening the robustness and application of skin cancer classification models, addressing these limitations and adopting these future approaches will also support the continued advancement of artificial intelligence in medical diagnostics. The comparative analysis, high accuracy, and creative approaches of this work serve as a strong basis for future studies that will improve and broaden the scope of skin cancer detection systems. Interestingly, any image classification domain data with classification use cases might be processed using the technique suggested in this study.

Funding Statement: The author(s) received no specific funding for this study.

Conflicts of Interest: There is no conflicts of interest regarding this study.

References

[1] P. C. Bizimana, Z. Zhang, A. H. Hounye, M. Asim, M. Hammad, and A. A. A. El-Latif, "Automated heart disease prediction using improved explainable learning-based technique," *Neural Comput. Appl.*, vol. 36, no. 26, pp. 1–30, 2024, doi: 10.1007/s00521-024-09967-6.

[2] H. Ullah, M. U. Haq, S. Khattak, G. Z. Khan, and Z. Mahmood, "A robust face recognition method for occluded and low-resolution images," in *Proc. 2019 Int. Conf. Appl. Eng. Math. (ICAEM)*, Aug. 2019, pp. 86–91, doi: 10.1109/ICAE.2019.8853753.

[3] M. U. Haq, M. A. J. Sethi, and A. U. Rehman, "Capsule network with its limitation, modification, and applications—A survey," *Mach. Learn. Knowl. Extract.*, vol. 5, no. 3, pp. 891–921, 2024.

[4] N. U. A. Tahir, Z. Zhang, M. Asim, J. Chen, and M. ELAffendi, "Object detection in autonomous vehicles under adverse weather: A review of traditional and deep learning approaches," *Algorithms*, vol. 17, no. 3, p. 103, Mar. 2023.

[5] D. Hemalatha, K. N. Latha, and P. M. Latha, "Skin cancer detection using deep learning technique," in *Proc. 2023 2nd Int. Conf. Innov. Technol. (INOCON)*, 2023, pp. 1–5.

[6] S. Gamil, F. Zeng, M. Alrifaey, M. Asim, and N. Ahmad, "An efficient AdaBoost algorithm for enhancing skin cancer detection and classification," *Algorithms*, vol. 17, no. 8, p. 353, Mar. 2023

[7] B. Ziyambe, A. Yahya, T. Mushiri, M. U. Tariq, Q. Abbas, M. Babar, et al., "A deep learning framework for the prediction and diagnosis of ovarian cancer in pre-and post-menopausal women," *Diagnostics*

(Basel), vol. 13, no. 10, p. 1703, May 11, 2023, doi: 10.3390/diagnostics13101703.
[8] M. Elgamal, "Automatic skin cancer images classification," *Int. J. Adv. Comput. Sci. Appl.*, vol. 4, no. 3, 2013, doi: 10.14569/IJACSA.2013.040342.
[9] The American Cancer Society, "Key Statistics for Melanoma Skin Cancer," *cancer.org*, 2022. [Online]. Available: https://www.cancer.org/content/dam/CRC/PDF/Public/8823.00.pdf. [Accessed: Dec. 30, 2024].
[10] M. Q. Khan, A. Hussain, S. U. Rehman, U. Khan, M. Maqsood, K. Mehmood, et al., "Classification of melanoma and nevus in digital images for diagnosis of skin cancer," *IEEE Access*, vol. 7, pp. 90132–90144, 2019, doi: 10.1109/ACCESS.2019.2926837.
[11] National Cancer Institute, "Skin cancer (including melanoma)—patient version," [Online]. Available: https://www.cancer.gov/types/skin. [Accessed: Sep. 17, 2021].
[12] Skin Cancer Foundation, "Skin cancer facts & statistics," Mar. 11, 2019. [Online]. Available: https://www.skincancer.org/skin-cancer-information/skincancer-facts/. [Accessed: Sep. 22, 2021].
[13] P. S. Staff, "What is the difference between melanoma and non-melanoma skin cancer?" *PSS*. [Online]. Available: https://www.premiersurgical.com/01/whats-the-difference-between-melanoma-and-non-melanoma-skin-cancer/. [Accessed: Dec. 30, 2024].
[14] R. R. Murphy, *Introduction to AI robotics*. Cambridge, MA, USA: MIT Press, 2019.
[15] A. Farag, H. R. Le Lu, J. Roth, E. Liu, O. Turkbey, and R. M. Summers, "A bottom-up approach for pancreas segmentation using cascaded superpixels and (deep) image patch labeling," *IEEE Trans. Image Process.*, vol. 26, no. 1, pp. 386–399, Jan. 2017, doi: 10.1109/TIP.2016.2624198.
[16] D. Bisla, A. Choromanska, R. S. Berman, J. A. Stein, and D. Polsky, "Towards automated melanoma detection with deep learning: Data purification and augmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2019, pp. 0–0.
[17] H. Rashid, M. A. Tanveer, and H. A. Khan, "Skin lesion classification using GAN-based data augmentation," in *Proc. 41st Annual Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 916–919, doi: 10.1109/EMBC.2019.8857905.
[18] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.
[19] L. Yu, H. Chen, Q. Dou, J. Qin, and P. A. Heng, "Automated melanoma recognition in dermoscopy images via very deep residual networks," *IEEE Trans. Med. Imaging*, vol. 36, no. 4, pp. 994–1004, Apr. 2017, doi: 10.1109/TMI.2016.2642839.
[20] H. A. Haenssle, C. Fink, R. Schneiderbauer, F. Toberer, T. Buhl, A. Blum, et al.; Reader Study Level-I and Level-II Groups, "Man against machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists," *Ann. Oncol.*, vol. 29, no. 8, pp. 1836–1842, Aug. 1, 2018, doi: 10.1093/annonc/mdy166.
[21] U. O. Dorj, K. K. Lee, J. Y. Choi, and M. Lee, "The skin cancer classification using deep convolutional neural network," *Multimedia Tools Appl.*, vol. 77, no. 8, pp. 9909–9924, 2018, doi: 10.1007/s11042-018-5714-1.
[22] F. Xie, H. Fan, Y. Li, Z. Jiang, R. Meng, and A. Bovik, "Melanoma classification on dermoscopy images using a neural network ensemble model," *IEEE Trans. Med. Imaging*, vol. 36, no. 3, pp. 849–858, Mar. 2017, doi: 10.1109/TMI.2016.2633551.
[23] F. Dalila, A. Zohra, K. Reda, and C. Hocine, "Segmentation and classification of melanoma and benign skin lesions," *Optik (Stuttg.)*, vol. 140, pp. 749–761, 2017, doi: 10.1016/j.ijleo.2017.04.084.
[24] A. R. Jadhav, A. G. Ghontale, and V. K. Shrivastava, "Segmentation and border detection of melanoma lesions using convolutional neural network and SVM," in *Computational Intelligence: Theories, Applications and Future Directions-Volume I: ICCI-2017*, Springer, Singapore, 2019, pp. 97–108, doi: 10.1007/978-981-13-1132-1_8.
[25] J. Daghrir, L. Tlig, M. Bouchouicha, and M. Sayadi, "Melanoma skin cancer detection using deep learning and classical machine learning techniques: A hybrid approach," in *Proc. 5th Int. Conf. Adv. Technol. Signal Image Process. (ATSIP)*, Sep. 2020, pp. 1–5, doi: 10.1109/ATSIP49331.2020.9231544.
[26] Vidya M and Karki M., "Skin cancer detection using machine learning techniques," in *Proc. 2020 IEEE Int. Conf. Electronics, Comput. Commun. Technol. (CONECCT)*, Jul. 2, 2020, pp. 1–5, doi: 10.1109/CONECCT50063.2020.9198489.
[27] J. A. Jaleel, S. Salim, and R. Aswin, "Artificial neural network based detection of skin cancer," *Int. J. Adv. Res. Electr. Electron. Instrum. Eng.*, pp. 200–205, 2012.
[28] S. Choudhari and S. Biday, "Artificial neural network for skin cancer detection," *Int. J. Emerging Trends & Technol. Comput. Sci.*, vol. 3, no. 5, pp. 147–153, 2014.
[29] C. Noel et al., "Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the International Skin Imaging Collaboration (ISIC)," *arXiv*, 2018. [Online]. Available: https://arxiv.org/abs/1902.03368.
[30] A. K. Waweru, K. Ahmed, Y. Miao, and P. Kawan, "Deep learning in skin lesion analysis towards cancer detection," in *Proc. 2020 24th Int. Conf. Information Visualisation (IV)*, Sep. 2020, pp. 740–745, doi: 10.1109/IV51561.2020.00130.
[31] Y. Huo, "Full-stack application of skin cancer diagnosis based on CNN model," in *Proc. 2021 IEEE Int. Conf. Comput. Sci., Electronic Inf. Eng. Intell. Control Technol. (CEI)*, Sep. 2021, pp. 754–758, doi: 10.1109/CEI52496.2021.9574583.
[32] D. Moldovan, "Transfer learning based method for two-step skin cancer images classification," in *Proc. 2019 E-Health Bioengineering Conf. (EHB)*, Nov. 2019, pp. 1–4, doi: 10.1109/EHB47216.2019.8970067.
[33] Y. Aydin, "A comparative analysis of skin cancer detection applications using histogram-based local descriptors," *Diagnostics (Basel)*, vol. 13, no. 19, p. 3142, Oct. 6, 2023, doi: 10.3390/diagnostics13193142.
[34] M. Shah, "LRNet: Skin cancer classification using low-resolution images," in *Proc. 2021 Int. Conf. Commun. Inf. Comput. Technol. (ICCICT)*, 2021, pp. 1–5.
[35] A. Tajerian, M. Kazemian, M. Tajerian, and A. Akhavan Malayeri, "Design and validation of a new machine learning-based diagnostic tool for the differentiation of dermatoscopic skin cancer images," *PLoS One*, vol. 18, no. 4, p. e0284437, Jun. 2021.