IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Mathematical Modelling of Engineering Problems
Vol. 12, No. 1, January, 2025, pp. 166-180
Journal homepage: http://iieta.org/journals/mmep

A Comparative Study of Deep Learning Models for Skin Cancer Detection: Leveraging Transfer Learning

Hassan Fakhry Hassan[1*], Sibel Tariyan Ozer[2]

[1] Department of Computer Engineering, University of Mosul, Mosul 41001, Iraq
[2] Department of Computer Engineering, University of Medipol, Ankara 06000, Turkey

Corresponding Author Email: Hasan.allayla@uomosul.edu.iq

https://doi.org/10.18280/mmep.120119

Received: 30 January 2024
Revised: 8 May 2024
Accepted: 15 May 2024
Available online: 25 January 2025

Keywords:
skin cancer detection, deep learning, dermatological images, Convolutional Neural Network (CNN), Xception, VGG16, NasNetMobile InceptionV3

ABSTRACT

Skin cancer is a significant global health concern, and early detection is crucial for successful treatment outcomes. Traditional diagnostic methods require expensive tools, such as a dermatoscope, which may not be available at all facilities. Artificial intelligence (AI) is one potential solution, which could provide accurate analyses of regular photos and determine whether any skin lesion can be found. The aim of this study is to develop AI-based models for skin cancer detection through the help of sophisticated deep learning (DL) algorithms. For evaluated models, including Convolutional Neural Network (CNN), Xception, InceptionV3, NASNetMobile, and VGG16. Note that their comparison metrics consist of accuracy, precision, recall values, F1 score, AUC on ROC curve, and specificity. The CNN model performed best, where it gave an accuracy of 97% and AUC 0.9962. The Xception model, with an accuracy of 95% and AUC of 0.9800, came next. Our InceptionV3 model played hard with an accuracy of 93% and AUC is 0.9819. Other models such as NASNetMobile and VGG16 showed less performance compared to the above model as follows with accuracy: 69%, AUC: 0.8228, and accuracy of 77%, AUC: 0.8057 correspondingly. Our study shows the promise of AI for increasing both the accuracy and availability of skin cancer diagnosis, opening up a much-needed addition in the fight to lower treatment-critical oncologic morbidity and mortality. Our work intends to create a superior model for skin cancer detection and improve medical image analysis and diagnostic tools.

1. INTRODUCTION

Skin cancer represents a great concern for global health expressed in a number of more than 3.5 million diagnosed cases per year, which include melanoma, basal cell carcinoma, and squamous cell carcinoma. An astounding factually presented is that this exceeds cancers such as breast, lung, and colon combined and that melanoma kills one person a minute [1]. Thus, the burden of SC underlines the importance of early detection through materialized and efficient treatment. Catching the disease early is crucial for successful treatment. For instance, the ten-year survival rate of detected cases is 94% if diagnosed early; if the cancer progressed, as in case melanoma, into stage four that rate drops to 15% [2].

This fact has inevitable practical implications – numerous lives could be saved, and a substantial amount of expensive advanced care-related costs could be eliminated. However, early SC detection is problematic in terms of it being expensive. Skin lesions are often visually similar, which results in a difficulty to differentiate between benign and malignant sources. Therefore, a comprehensive examination is to be conduct. Traditionally, dermatologist practitioners utilize a medical device called a dermatoscope to record images of the lesion to take a closer look. Unfortunately, the tool is expensive, and not all practitioners have access to it, which stands as an obstacle in the way of cheap and effective early detection. Skin diseases' appearance is visually difficult, as conditions look similar.

However, there is a way, which was develop due to current advancements in AI. Modern algorithms are capable of identifying visually similar skin conditions without the additional need for a medical tool. Instead, based on this algorithm, usual photos can be analyzed. The implementation of artificial intelligence in the context of SC detection in the healthcare sector was exceptionally promising. SC detection accuracy was increase by DL, which allows the assessment of complex patterns & features of dermatological images [3]. AI algorithms to infer potential signs of SC could carefully analyze skin images. In many cases, these systems demonstrate a sufficiently high ACC and may offer healthcare workers a second opinion, highlight fields of cold interest in the specific pictures, and explain potential diagnoses. Given the rapidly increasing spread of skin cancer around the globe and the urgent need for accurate diagnostics, this application of ACC and AI is highly applicable.

The focus of this research is to implement the thorough process of SC detection through image data analysis. The journey commences with accurate data processing, with the

outline including the essential procedures of pre-processing and cleaning. Next is the dataset division into separate train and test sets, which paves the way for sustainable model training and evaluation.

Subsequently, we explore data modeling, using advanced methodologies to identify potentially interesting patterns and characteristics in the images.

Utilizing the above-mentioned algorithms and models, as well as making some critical changes in the hyperparameter tuning process, we aspire to develop a powerful and accurate model for SC detection. This work aims to attain not solely the best performance but also to contribute to other areas of medical image analysis and diagnostic tools.

2. SKIN DISORDERS MEDICAL

The skin, documented as the body's largest organ [4], plays a multifaceted role in maintaining overall health. It serves to regulate body temperature, provides protection against injuries and infections, contributes to vitamin D production, and functions as a reservoir for fat and water storage. Structurally, the skin comprises three primary layers: the outermost Epidermis, the inner Dermis, and the deepest layer, Hypodermis. Skin disorders, encompassing conditions that impact these layers [5], manifest in various ways, leading to symptoms such as rashes, sores, itching, or other observable changes. The origins of skin disorders are diverse, with some linked to lifestyle factors and others rooted in genetic predispositions. The spectrum of skin disorders ranges widely in severity and symptoms, encompassing genetic or situational causes, permanence or temporality, painless or painful manifestations, and variations in life-threatening potential [6]. As seen in Figure 1, some types of cancer could be seen clearly. Melanoma stands out as one of the most life-threatening cancers globally, capable of metastasizing to other organs.

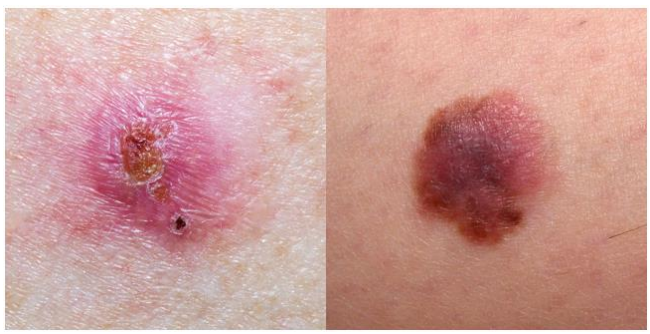

Figure 1. Skin cancer overview [7]

If not detected and addressed in its early stages. The subsequent sections will delve into the progression of melanoma, pertinent statistics, and additional information associated with this malignancy [5].

Situated directly above the dermis, the uppermost layer of the epidermis harbors melanocytes, responsible for skin pigmentation. However, when these melanocytes undergo uncontrolled growth, a malignant tumor, known as melanoma, emerges [8, 9]. This tumor possesses the potential to spread to various parts of the body. There are instances where pre-existing normal moles or nevi on the skin transform into melanomas, marked by observable changes such as alterations in shape, border, size, or color [10]. Melanoma predominantly manifests in regions like the scalp, face, trunk (abdomen, back, and chest), arms, and legs, though it can also occur in less sun-exposed areas, including the neck and head. Cutaneous melanoma, initiating in the skin, is the most prevalent type, characterized by three main subtypes. The first, "superficial spreading melanoma", arises in up to of melanoma cases and often originates from an existing mole. The second type, "lentigo maligna melanoma", tends to affect older individuals, typically emerging in sun-exposed regions. Accounting for about of melanomas, the third type is "nodular melanoma", typically identified as a raised bump on the skin [11].

3. RELATED WORKS

Considerable research has focused on the classification of skin lesions from dermoscopic images, but limited work been conducted on classification using general images.

A notable research effort led by Esteva and colleagues [2] has emerged as one of the most comprehensive studies skin lesion classifications. They carefully classified lesions into 23 distinct categories, utilizing CNNs for the job. Specifically, they utilized the capabilities of VGG-16 and VGG-19 models [12]. Esteva and the team enhanced the performance of a pre-trained VGG model through transfer learning. Notably, they accomplished an impressive 90%.

ACC rate in binary classification, successfully distinguishing between cancerous and non-cancerous lesions.

In the field of melanoma detection, Moussa and colleagues adopted a unique approach, identifying the condition based on its geometric features. Moussa et al. [13] employed the k-Nearest Neighbors algorithm for classification. Despite the constraint of a limited dataset, their method demonstrated an impressive 89% ACC rate.

Alternative approaches have investigated the use of a Total Dermatoscopy Score (TDS) to distinguish between malignant and benign skin lesions. Azmi et al. [14] explored this method, assigning TDS scores by extracting features following the ABCD rule. The scores varied from 1.0 to 8.9, with a score above 5.45 indicating an elevated likelihood of melanoma.

Masood and collaborators led a research project as conducted by Masood et al. [15]. In this investigation, a set of 135 images underwent classification into cancerous and non-cancerous lesions using an Artificial Neural Network (ANN) system. The methodology involved the application of the Histogram Analysis-based Fuzzy C Mean Algorithm for Level Set Initialization to segment the images. Following the segmentation process, features were extracted, encompassing histogram features and statistical features derived from the Grey Level Co-occurrence Matrix (GLCM).

They employed a two-layer feed-forward neural network for classification, incorporating three distinct training algorithms: Levenberg-Marquardt, Scaled-Conjugate Gradient, and Resilient Back Propagation. Remarkably, the study attained an accuracy rate of 91.9% when utilizing the Scaled Conjugate Gradient training algorithm.

Many researchers have actively explored the computer vision approach for detecting SC. In the segmentation of skin lesions in input images, current systems often use manual, semi-automatic, or fully automatic border detection methods. These segmentation techniques depend on diverse features, such as texture, color, shape and luminance. The literature includes numerous border detection methods, including histogram thresholding [16], global thresholding on optimized color channels followed by morphological operations, and hybrid thresholding [17].

In our study, we applied an automatic thresholding and border detection method, utilizing various image-processing techniques to extract these features. For example, Garnavi [18] introduced an automated method for global border detection in dermoscopy images, relying on color-space analysis and global histogram thresholding. This method demonstrated high performance in detecting the borders of melanoma lesions. As reported by Celebi et al. [19], the approach included segmenting the input image into different clinically significant regions using the Euclidean distance transform to extract color and texture features.

The ABCD rule of dermoscopy emphasizes asymmetry as a fundamental feature, alongside border irregularity, Color, and diameter. Numerous studies have aimed to quantify asymmetry in skin lesions. Some approaches compute symmetry using geometric measurements across the entire lesion, incorporating metrics like symmetric distance and circularity [20]. Conversely, other research suggests the circularity index as a measure of border irregularity in dermoscopy images [21, 22].

Maglogiannis and Doukas [23] presented an overview of the most notable implementations in the literature. They compare the performance of various classifiers for the specific problem of skin lesion diagnosis, providing valuable insights into the field of SC detection.

According to Shorfuzzaman's study [24], malignant melanoma, a highly fatal skin cancer originating from abnormal melanocyte proliferation, must be detected early for cure. In the present work, an explainable CNN-based stacked ensemble framework is suggested for automated melanoma diagnosis. Multiple CNN sub-models EfficientNetB0, Xception, and DenseNet121 are developed through transfer learning, and their forecasts are used by a separate meta-learner. The ensemble model was assessed using open-access data that included both benign and malignant melanoma images. An explainability strategy with Shapley adaptive explanations is used to generate heat maps that show the melanoma-suggestive zones, rendering the model more interpretable for dermatologists. The ensemble framework accomplishes the highest performance, reaching an accuracy of 95.76%, a sensitivity of 96.67%, and an AUC of 95.7%, surpassing individual models.

According to Tembhurne et al. [25], skin cancer is a prevalent disease, with melanoma being particularly deadly, causing six out of seven skin cancer-related deaths. To improve diagnosis accuracy and reduce false negatives, this paper presents a novel approach combining machine learning and deep learning techniques. The model uses neural networks for feature extraction and processes these features with methods like Contourlet Transform and Local Binary Pattern Histogram. Achieving an accuracy of 93% and recall scores of 99.7% for benign and 86% for malignant cases, the model was teste on a Kaggle dataset from the ISIC Archive. This ensemble method outperforms expert dermatologists and existing techniques, offering a valuable tool for accurate skin cancer detection.

Imran et al. [26] introduced an ensemble of deep learners, specifically VGG, CapsNet, and ResNet, for skin cancer detection. The proposed ensemble model outperforms individual models, achieving superior results in sensitivity, accuracy, specificity, F-score, and precision. The experimental results, with the proposed ensemble achieving an accuracy of 93.5%, sensitivity of 0.87, specificity of 0.84, F-score of 0.92, false-positive rate of 0.06, and precision of 0.94, demonstrate the effectiveness of this approach and suggest its potential application to other disease detection tasks. In Table 1, there are overview description of related works.

Table 1. Overview of related works in skin cancer detection

Study	Approach	Key Techniques	Dataset	Performance
Esteva et al. [2]	Skin lesion classification	VGG-16, VGG-19, Transfer Learning	Multiple datasets	Accuracy: 90%
Moussa et al. [13]	Melanoma detection	k-Nearest Neighbors, Geometric Features	Limited dataset	Accuracy: 89%
Azmi et al. [14]	Malignant vs. benign skin lesions	Total Dermatoscopy Score (TDS), ABCD rule	N/A	TDS score > 5.45 indicates high likelihood of melanoma
Masood et al. [15]	Cancerous vs. non-cancerous lesions	Artificial Neural Network (ANN), Histogram Analysis-based Fuzzy C Mean Algorithm, Grey Level Co-occurrence Matrix (GLCM)	Set of 135 images	Accuracy: 91.9%
Multiple studies [16-22]	Segmentation of skin lesions	Various border detection methods (histogram thresholding, global thresholding, morphological operations, hybrid thresholding)	N/A	High performance in detecting melanoma lesion borders
Maglogiannis and Doukas [23]	Overview of skin lesion diagnostic implementations	Comparative analysis of classifiers for skin lesion diagnosis	N/A	N/A
Shorfuzzaman [24]	Melanoma detection	EfficientNetB0, Xception, DenseNet121, Transfer Learning, Shapley adaptive explanations	Open-access dataset	Accuracy: 95.76%, Sensitivity: 96.67%, AUC: 95.7%
Tembhurne et al. [25]	Skin cancer detection	Neural Networks, Contourlet Transform, Local Binary Pattern Histogram	Kaggle dataset (ISIC Archive)	Accuracy: 93%, Recall (Benign): 99.7%, Recall (Malignant): 86%
Imran et al. [26]	Skin cancer detection	VGG, CapsNet, ResNet	N/A	Accuracy: 93.5%, Sensitivity: 87%, Specificity: 84%, F-score: 92%, Precision: 94%
Vineeth et al. [27]	Early skin cancer detection	CNN with three hidden layers, Hybrid combination of activation functions	N/A	Accuracy: 95%
Guergueb and Akhloufi [28]	Melanoma detection	Convolutional Neural Networks	36,000 images from multiple datasets	Accuracy: > 99%, AUC: > 99%

Vineeth et al. [27] studied a serious public health problem associated with the growing prevalence of common skin cancer around the world. The authors used deep learning techniques to create a model capable of predicting the disease in its early stages. A CNN model, which was designed by applying images of skin to analyze cancer-related attributes, brings desirable results in three hidden layers using the hybrid combination of activation function results in an accuracy of 95%. The system predicted with high accuracy for the test data of unseen data, concluding that the system can be used as an effective platform for the early diagnosis of skin cancer to facilitate newcomers and assist doctors, doctors and clinics.

Guergueb and Akhloufi [28] focused on melanoma, one of the most lethal types of skin cancer, capable of metastases if detected too late. According to the paper, the author sought to investigate the application of state-of-the-art Convolutional Neural Networks to detect melanoma, involving more than 36,000 images extracted from several datasets. The outcomes indicated that the top-performing model based on deep learning delivers extraordinary classification accuracy and a high Area under the Curve exceeding 99%, which illustrates the significant contribution of the model to the advancement of melanoma detection and early examination.

Kadampur and Al Riyaee [29] solved the problem of early detection of skin cancer, through the developed deep learning-based models to classify dermal cell images. The study uses a cloud-based model-driven architecture combined with advanced deep learning-based algorithms to improve the prediction performance. The models are evaluated using standard datasets, and an AUC close to 99.77% was achieved. Besides, InceptionV3 obtains the highest precision of 98.19% with an AUC of 99.23%. The results show that the proposed model-driven architecture can assist the practitioners in developing highly accurate skin cancer prediction models.

4. METHODOLOGY

A systematic approach to our methodology on developing AI-based models for skin cancer detection. The process can be further broken down into multiple steps: data pre-processing, data splitting, modeling and evaluating [30]. The first step would be to get the dataset of skin cancer, which contain images of all types of skin lesions [31] after going through an extensive data preprocessing, at last my dataset finally ready for modelling, and Figure 2 describes general flowchart for the processing. Important preprocessing includes but not limited to normalizing, flattening and reshaping the data for input into model. Other things we may consider is SMOTE (Synthetic Minority Over-sampling Technique) [32] if our classes are imbalanced Preprocessed data was then split to training and test dataset using common train-test ratio (80% training and 20% testing) in order to have enough data for the model learning, but also reliable evaluation. After preprocessing and splitting, the same data was used to train a number of advanced deep learning models.

Convolutional Neural Networks (CNN): The architecture that we use to analyze an image is the one from CNNs because it is impressive in capturing spatial hierarchies in other images using convolutional layers. InceptionV3, a state-of-the-art classification network which used multiple filters of different sizes for convolutions to be able to capture information at various scales. Xception, an extreme version of separable depthwise convolutions improved parameter efficiency with experimental evidence that it performs better than similar deep networks at lesser computational cost and faster execution time: NASNetMobile developed for mobile-devices expected less powerful machines with low powered computation resources and VGG16 pretrained CNN widely recognized high performance on object detection tasks.

For each model, we train on the training set and validate on the validation set. The final step is to evaluate the performance of all models, using a range of metrics. Those are accuracy, precision, recall or, F1-score, specificity, and AUC. This systematic method ensures that the models are trained and tested in a controlled manner, utilizing the power of state-of-the-art deep learning architectures to enhance both predictive performance and availability for skin cancer detection.

4.1 Data description

The dataset used in this study was obtain from the ISIC Archive. ISIC is the International Skin Image Collaboration and ISIC Archive is a well-known source of dermatological images that supports research regarding the diagnosis of skin cancer. The dataset comprises 3,297 skin lesion images; specifically, it contains 1,800 benign and 1,497 malignant cases as shown in Table 2.

All the images are resized to 224×224 pixels and are in RGB format to ensure that the data is consistent for model training and testing. The main purpose of using this dataset is to develop a machine-learning model that can distinguish between two classes: benign and malignant skin lesions [33].

It is inherently an extremely difficult task for several reasons discussed earlier paper. Firstly, skin inflammation lesions vary greatly in color and texture, while some even have spiky or asymmetrical borders. Such characteristics are often difficult, even for experienced human practitioners to differentiate, especially in the early stages when a skin cancer lesion can resemble a mole.

The dataset image resources are linked to the secondary aim of creating an improved skin cancer detection algorithm [34]. Early detection is a key factor in patient survival. By developing advanced image analysis and classification tools, we can automate this process, making it less tiring for doctors, cost-effective and, more importantly, increase accuracy. The visual example from the dataset, accompanying this guide, highlights the multiple complexities in determining whether primary skin cancer is benign or malignant. The example images demonstrate the need for models that can easily identify these subtle variations.

Hence, the dataset, comprising high-quality images of moles and rashes, is comprehensive and can be advantageous in developing and accessing deep learning models.

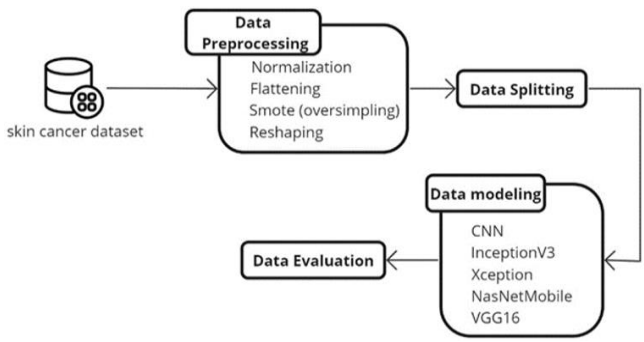

Figure 2. General flowchart for skin cancer detection

Table 1. Description of the dataset

Attribute	Description
Source	ISIC (International Skin Image Collaboration) Archive
Total images	3,297
Benign images	1,800
Malignant images	1,497
Image resolution	224×224 pixels
Image format	RGB
Classes	Benign, Malignant
Objective	Develop a deep learning model for classifying skin lesions into benign and malignant categories
Challenges	Intricate and nuanced variability in skin lesion appearances, making visual classification difficult
Application	Enhance early detection of skin cancer, assist dermatologists in clinical settings
Key features	Color, texture, shape variability in images
Significance	Aids in improving diagnostic accuracy and timely intervention for skin cancer patients

4.2 Data splitting

In the realm of ML and data analytics, the foundational practice of segmenting a dataset into distinct subsets [35] paramount. This segmentation typically results in the creation of a training set and a test set. The training set plays crucial role in sculpting the ML model, shaping its understanding and capabilities. On the other hand, the test set serves as a litmus test for the model's performance on unfamiliar data, evaluating its ability to generalize and guarding against over fitting. In our methodology, we strictly adhered to this established practice, allocating 20% of the dataset specifically for testing purposes.

4.2.1 Hyperparameter tuning and the training process

In our methodology, extensive hyperparameter tuning was performed using a grid search strategy to determine the best settings for each convolutional neural network model, including InceptionV3, NASNetMobile, VGG16, and Xception. The strategy allowed us to comprehensively evaluate and select optimal hyperparameters for the learning rate, batch size, and number of training epochs [36], which helped to balance diagnostic accuracy and computational efficiency. The models were trained using extensive computational resources to efficiently process the vast amounts of dermatological image data and demands on the neural network architecture to achieve a well-suited diagnostic model. The high-performance computing environments were crucial for the training process based on the huge computations and time required to train a model on such datasets. Continuous validation checks were performed throughout the training to track the model's performance on unseen data and to avoid overfitting to ensure that the model generalizes well in new clinical environments. Utilizing validation sets during the training phase enabled us to fine-tune the model to achieve the best performance without biased data compared to the model training set, and Figure 3 views samples for the dataset. Furthermore, the computational resources were used to apply advanced data augmentation techniques to display the input images using random rotations, shifts, zooms and flips to produce robustly trained models that can predict extensive variations observed on new clinical images with such variations. The training setup on the computational environment allowed for significantly faster computations and a more robust and reliable diagnostic model that could achieve high efficiency in real-world clinical settings.

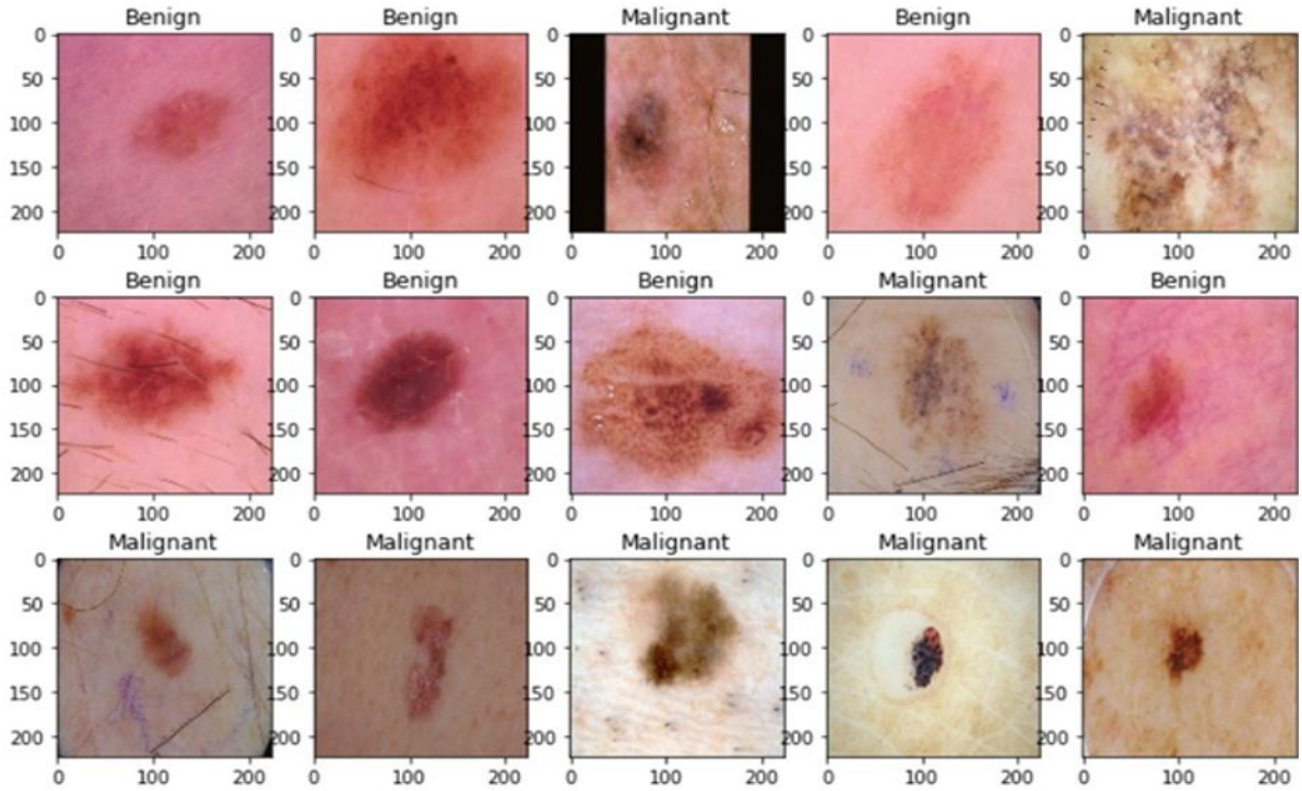

Figure 3. Overview of the dataset

4.3 Data modeling

Data modeling is another critical factor in SC detection, enabling the creation of an accurate and reliable system for early diagnosis of the disease. It is achieved through the use of multiple datasets of skin images and implies the development of computational models capable of identifying patterns signaling benign or malignant lesions. As a data modeling is another critical factor in SC detection, enabling the creation of an accurate and reliable system for early diagnosis of the disease. It is achieved through the use of multiple datasets of skin images and implies the development of computational models capable of identifying patterns signaling benign or malignant lesions.

4.3.1 CNN model

Convolutional Neural Networks are particularly good at processing structured grid information, such as images where they can directly learn features from the relevant pixel values. In the image recognition sense, CNNs will use its convolutional layers to identify patterns and spatial relationships in the relevant inputs, pooling layers to create compressed representations and fully connected layers to enable classification. In SC detection, the task involves identifying patterns in dermatological images [36], where CNNs are good at distinguishing features from a well-trained set. Here, diversity in data sources while training a CNN allows the created model to discern the differentiation point between the benign and malignant cases from the unique visual features, enhanced by the hierarchical model's ability to capture local and global patterns. Notes for my visualization-the CNN architecture involves:

- Convolutional Layers apply convolution operations to the input images using filters to detect features such as edges, textures, and patterns. The filter produces a feature map that highlights aspects of the input that are useful.
- Activation Functions like ReLU introduce nonlinearity into the network to learn complex patterns. We then have the Pooling Layers like max-pooling, which is used to reduce the spatial dimensionality.
- Fully Connected Layers then perform high-level reasoning by connecting all the previous layers' neurons to every other neuron creating a classification. The exploration and learning of these features from the images create an excellent visual identification model such as CNN for the SC examples because of varied skin lesion features.

4.3.2 InceptionV3

InceptionV3 [37] is a state-of-the-art Convolutional Neural Network architecture developed for image classification [38] and object recognition tasks. Building on the success of the predecessors, InceptionV1 and InceptionV2, InceptionV3 includes several innovations that significantly improve its performance and efficiency. One of the defining features of InceptionV3 is the use of inception modules, specifically designed to capture useful features at many scales.

These modules utilize filters of different sizes in the same level, enabling the network to simultaneously learn and recognize patterns of different granularities. The use of inception modules helps address the problem of overfitting and reduce computational cost, making efficient and parallel work with each dimension of input data. For SC detection, InceptionV3 is very capable of recognizing subtle patterns in dermatological images. Even if we consider benign and malignant lesions, we can notice slight changes in the lesion's pattern.

InceptionV3's architecture helps detect this kind of slight differences, making the prediction of skin abnormalities accurate. InceptionV3 uses complicated architecture and deeper learning, meaning it can extract vital features in dermatological images that it was trained on. These pre-trained weights then lay a solid foundation that allows us to fine-tune the model on the new dataset, which is the skin lesion image dataset in our case.

By fine-tuning InceptionV3 on the skin lesion image dataset, we can adapt the learned features to the skin lesion's characteristics, thus increasing the accuracy of classification. There are local and global features that can be extracted from the image, and extracting both local and global features is important for SC detection. Local features include small color variations and minor textural changes that are critical for detecting SC in the early stage.

4.3.3 NASNetMobile

NASNetMobile is a lightweight architecture designed for low-compute environments [39, 40], such as mobile or embedded devices. The method of its creation, neural architecture search, is based on the execution of multiple trials of machine learning models, resulting in high-performing yet efficient architectures. The key highlight of neural abstraction inquiry behind the creation of system's architecture is the ability to provide an automated method for discovering which structures and arrangements are most efficient for task completion. Therefore, NASNetMobile is highly beneficial for skin cancer detection, as it offers a combination of the ability to handle processing of complex data patterns while using fewer resources to do so. In other words, it features sufficient light-weightedness to be deployed on devices with limited memory, which is a major issue for deep-learning models. However, its light structuring does not imply inferior performance – on the contrary, it appears to be robust when it comes to the accurate classification of skin lesions. From a practical perspective, deploying a skin cancer detection model trained on NASNetMobile directly on a mobile device, it can work under the constraints of mobile computing power memory. It simply means that such a model can provide real-time labels for skin images without any need in internet connection or remote server. Meanwhile, with more data provided from the user, the model training on NASNetMobile may become more proficient in recognizing patterns in dermatological images. This fine-tuning will include adjustment to the existing model parameters to fit the particularities of skin images – their irregular patterns, diverse textures, and color distinctions of malignancy vs. benignancy. To sum up, NASNetMobile was chosen as the basis for the skin cancer detection system due to its ability to provide high performance and low computational cost. These are the criterion features for the application designed to run on mobile device, helping physicians in classifying skin lesions with high accessibility and ubiquity required by the real-world application.

4.3.4 VGG16

There is no such thing as "VGG16 Mobile" as a standard model; however, the original VGG16 architecture, which is famous for its high performance on image classification tasks, can be adapted for utilization in mobile applications. VGG16

[41] was the original network model with 16 layers; it has been one of the most successful deep convolutional neural networks for image recognition in in a variety of tasks. The simplicity but depth of the network architecture makes complex patterns in image data detectable. How may I adapt VGG16 for mobile? In order to make it feasible to run a neural network on a mobile device, model optimizations may be carried out with several adjustments that would lower the computational load and the resources required to classify with high accuracy. The following optimizations are possible:

$$P = \frac{TP}{TP + FP} \tag{1}$$

- Pruning – remove less important connections in the network to minimize the number of parameters and the computational load.
- Quantization – reduce the number of units in numerical parameters to reduce the size of the model, making inference faster without losing much of the performance.
- Knowledge distillation – train a smaller model to learn from the output of a larger model.

Overall, these techniques can be combined to achieve creating a compressed version of VGG16's architecture that is still capable of extracting deep features out of images, which is vital for the accurate analysis of dermatological images. Fine-tuning this model with a skin lesion dataset is essential in enabling the model to detect the intricate patterns behind each type of lesion, including benign and malignant tumors, by adjusting the network's weights in a way that increases its sensitivity and specificity in distinguishing between skin conditions specifically, which is essential for accuracy in clinical settings. The rationale for keeping VGG16 for the task of deploying it on mobile devices SC diagnosis centers is that this model has been observed to perform well in image classification and that, despite its depth, it can be effectively compressed for the challenges of mobile deployment. This is because the layers in this model are structured such that they capture the entire range of features in the images, which is suitable for mobile deployment mostly when the application is medical and accuracy is crucial. Thus, compressing VGG16 strikes a balance between keeping a deep network for robust analysis while compressing it enough to enable reliable performance on mobile devices, making it an ideal choice for such deployment. In conclusion, VGG16 can be compressed to meet the requirements of mobile deployment while maintaining image analysis capabilities for detecting skin cancer diagnosis. It would promote more advanced dermatological diagnostics deployment on mobile platforms and help patients diagnose their problem earlier.

4.3.5 Xception

The Xception model, developed by Google, is another sophisticated CNN that uses an innovative approach to known convolutional layers. Xception uses what are called depth-wise separable convolutions, which is a slight but essential deviation from the typical convolutional layer topology [42]. This modification allows the model to work more effectively with parameters and computational resources, which is important for processing large volumes of high-resolution images, such as dermatological images used in SC applications. With depth-wise separable convolutions, the Xception model processes layers separately along the spatial and depth dimension. This design significantly decreases the cost of computation and preserves the model's capability to capture complex image elements. The latter quality is critical, as dermatological images often rely on subtle message that differentiate benign patterns from malignant ones [43].

$$Accuracy = \frac{Correct\ Prediction}{Total\ data\ points} \tag{2}$$

Thus, fine-tuning Xception with a larger and diverse database of skin images is a crucial step in ensuring that the model can be used to detect SC. Investing more images into training the model in the database of dermatological images that exists during fine-tuning enables it to generalize across a much broader range of different pictures. Xception can then identify a wide variety of textures, shapes, and colored patches unique to different kinds of potentially dangerous skin routines, increasing its diagnostic accuracy. High precision and performance are potentially beneficial for integrating Xception into algorithms for automated SC diagnosis. Moreover, the ability to handle very complex skin images with a large number of fewer parameters than conventional CNN is a very attractive quality, especially for domains such as medicine, where both precision and efficiency are essential [43].

In brief, the Xception model's depth and the novel approach to conversation process make it suited to SC diagnosis needs. The ability of depthwise separable convolutions enables a high better performing model with much less computational demand. This would be crucial whenever the computational unit is limited, or the higher away demands for high-resolution images, which are processed very fast. The application of Xception in SC diagnosis will improve diagnostic outcomes and make health care more accessible and efficient for the SC screening process.

4.4 Evaluation metrics

When assessing our skin cancer (SC) detection model, we go beyond mere accuracy (ACC). Precision (PREC), Recall (REC), and the F-measure give us a fuller picture by detailing the model's precision, its ability to identify all actual cases of SC, and a balanced score of both, respectively. We also utilize the confusion matrix (CM) to examine the model's predictions across different classes [44, 45]. This comprehensive evaluation strategy ensures a more nuanced understanding of the model's effectiveness in SC detection

4.4.1 Accuracy

ACC, as outlined by Gu et al. [46] in their study on evaluation metrics, is a measure used to determine the effectiveness of a classifier in making correct predictions. It quantifies this proficiency by evaluating the count of correct predictions in relation to the total predictions made. The calculation for ACC is expressed as follows:

4.4.2 Precision

PREC, as cited in the work by Kynkäänniemi et al. [47], assesses a classifier's PREC by calculating the ratio of true positive (TP) to the sum of TP and FP. A higher PREC value signifies greater PREC with fewer FP, while a lower value implies the opposite. Essentially, PREC quantifies the ACC of a classifier based on its capacity to minimize misclassifications. The formula for calculating this metric is as follows:

$$P = \frac{TP}{TP + FP} \tag{3}$$

4.4.3 Recall

As given in the study of Grandini et al. [44], REC quantifies whether a classifier can correctly recognize all the positive instances; thus, it demonstrates that the sensitivity is reflected. A higher REC score means there are fewer positive instances that were missed, which translates to lower false negative (FN) values. In essence, REC equals the ratio of TP to the summation between TP and FN. This metric is calculated as:

$$R = \frac{TP}{TP + FN} \tag{4}$$

4.4.4 FI-Score

The F1-score (F1-S) [48], as you can see in the formula below, is a composite measure that combines both PREC and REC by using their harmonic mean with weight: This metric can be thought of providing an overall measure towards our models ability to predict PREC and REC. This can be written out as the formula to calculate F1-S.

$$\text{F1 - score} = 2\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

5. MODEL TRAINING AND EVALUATION

In the research methodology, we used Collab pro which is a cloud-based computing platform that helped us in doing our work more efficiently as well as effectively.

The high-performance computational environment, equipped with powerful GPUs, enabled the efficient training of machine learning models, overcoming the limitations commonly associated with standard hardware infrastructures [49]. Collab Pro enabled fast model training and validation due to the considerable amount of local configurations that were not needed [50]. In addition to this, the collaboration features enabled us to distribute code, share data and communicate with other members in our research team [51], creating a conducive work environment. All in all, Collab Pro helped us focus on conducting our SC detection experiment resulting into a seamless research experience. The current methodology is based on the application of cutting-edge convolutional neural network architectures: InceptionV3, NASNetMobile, VGG16, and Xception. These models are internationally recognized for their high efficacy in dealing with complex image data. The multi-model basis of the current study allows the integration of the most effective CNN architectures and ensures that the method is scalable and generalizable. For example, although Esteva et al. [2] used VGG models to achieve high classification accuracy of skin lesions, the current methodology applies more than a single model and enhances their performance by fine-tuning them to better process images of skin lesions. This approach not only improves the array of anomalies that can be detected via the models but also reduces the risk of overfitting, which is common in the single-model method. The methodology also operates a thorough pre-processing routine to ensure that the images fed to the neural networks are of high quality, increasing training and performance accuracy. In the study conducted by Moussa et al. [13], detection of melanoma relied on geometric features, making its analysis less informative than the current approach. The systematic use of advanced evaluations as accuracy, precision, and confusion matrices permits a detailed evaluation of each model's advantages and alignment with practical use in clinical practice. Thus, not only does the current research follow the recent trends in medical imaging development, but it also presents a novel, versatile, and scalable framework that exceeds many established approaches in the field of skin cancer detection.

5.1 CNN model

We used strict statistical validation applications to assure the robustness and reliability of performance related to the CNN model utilized in skin cancer detection in this study. We used k-fold cross-validation to divide the dataset onto five distinct subsets to ensure that each subset is used as a testing set in one of the five folds, while the other four categories are used to train the model. This method eliminates bias and variance hence giving a reliable measure of how the model can perform across different subsets of the data used.

Our CNN model for skin cancer detection demonstrates exceptional diagnostic capabilities, achieving an overall accuracy of 97%. This high accuracy is crucial in clinical settings, where correct classification significantly impacts patient outcomes. Table 3 views the classification report for the database of CNN model. The model's sensitivity, or true positive rate, is 100%, ensuring nearly all malignant lesions are identified, which is vital for early detection and treatment of melanoma.

The specificity, at 94.44%, indicates the model's effectiveness in identifying benign lesions, although there is a slight inclination towards false positives. Balancing sensitivity and specificity are essential for clinical usability, aiming to reduce false positives without compromising the identification of malignant cases.

Table 2. Classification report of CNN model

	Precision	Recall	F1-Score
0	1.00	0.94	0.97
1	0.94	1.00	0.97
Accuracy		0.97	
Sensitivity		1.00	
Specificity		0.94	
Macro avg	0.97	0.97	0.97

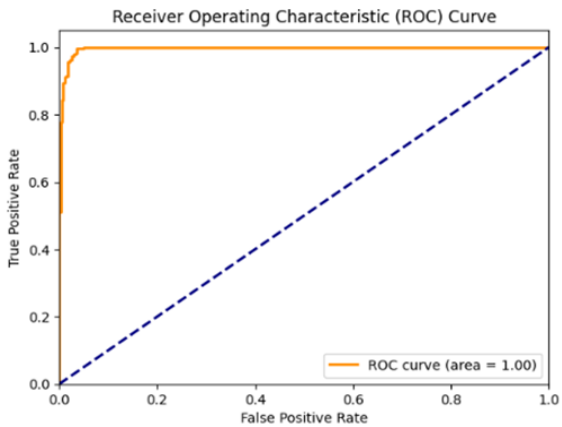

Figure 4. ROC of CNN

Figure 5. Loss vs. accuracy

Furthermore, the area under the Receiver Operating Characteristic curve of 0.996 is an indicator of a high discriminative ability. The ROC curve is a graphical representation of the diagnostic ability of a binary classifier system. Figure 4 shows that the model has high confidence in distinguishing between benign and malignant skin Furthermore, the area under the Receiver Operating Characteristic curve of 0.996 is an indicator of a high discriminative ability. The ROC curve is a graphical representation of the diagnostic ability of a binary classifier system.

Training and validation loss and accuracy curves provide deeper insights into the model's learning dynamics. The steady decrease in training loss coupled with consistent improvements in training accuracy demonstrate the model's effective learning from the training data. However, the observed fluctuations in the validation loss suggest moments of lesser generalization, which might be due to variations in data distribution or overfitting on the training data. These spikes highlight the importance of regularization techniques or potentially exploring more robust or diverse training data to enhance the model's generalization capabilities (Figure 5).

In summary, the CNN model is highly accurate and sensitive and specific and should be implemented in clinical settings for skin cancer detection. The ratio of sensitivity to specificity could be further improved, which could enhance the model's reliability.

5.2 InceptionV3

Our InceptionV3 model in skin cancer detection has shown excellent diagnostic performance which is confirmed from the Table 4. The model achieved an accuracy of 93% which shows high correctness of classification of skin lesions. Its sensitivity was 92.67% which indicates that this model can effectively identify all the malignant lesions which is essential in early diagnosis. The model's specificity was 93.33% which shows the models capacity to perform True Negative classification and then correctly identify the number of benign cases to help in the reduction of unnecessary intervention. The model's precision for the malignant class was 92% and that of the benign was 94%. This indicates that there is a high level of correct positive predictions and a low false positive for the benign class. The model's F1-score for the two classes was 0.93 which shows that the model can well balance the capacity to classify benign and malignant lesions.

Table 3. Classification report of InceptionV3 model

	Precision	Recall	F1-Score
0	0.94	0.93	0.94
1	0.92	0.93	0.92
Accuracy		0.93	
Sensitivity		0.92	
Specificity		0.93	
AUC-ROC		0.98	
Macro avg	0.93	0.93	0.93

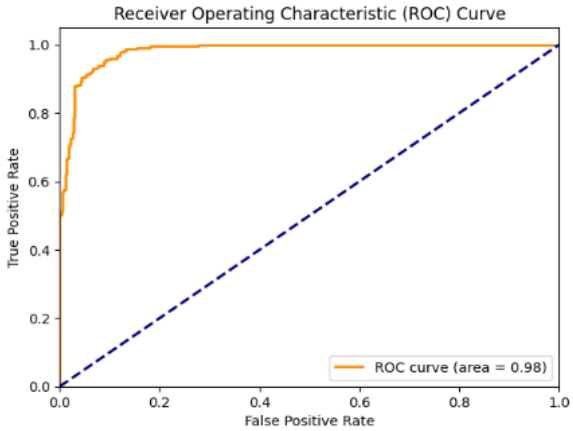

Figure 6. ROC of InceptionV3

The area under the ROC curve was 0.9818, which indicates the model's excellent discriminative ability. The high AUC value means that the model is extremely effective at distinguishing between benign and malignant cases. This is a crucial feature to assure the accuracy of diagnostic in clinical practice. Based on the principles of AUC, a value close to 1.0 indicates a high true positive rate with an extremely low false positive scale. Therefore, the model is reliable when it comes to accurately identifying potential skin cancer in the early stages. A_ROC curve visually confirms the model's performance. When representing the true positive rate against the false positive rate at different threshold settings, the curve explains sensitivity and specificity's trade-off. When the curve is closer to the upper left corner of the plot, it means that its performance is high. An AUC of 0.98 confirms the high model quality in terms of distinguishing benign and malignant cases (Figure 6).

The curve of training and validation loss decrease continuously indicating successful learning, and the accuracy curves show constant growth. Yet, the spikes of validation loss

are indicators of overfitting, emphasizing the necessity of regularisation and various training data to foster the capacity of generalization. Therefore, the InceptionV3 model shows strong performance in skin cancer detection, with high accuracy, sensitivity, specificity, and great AUC (Figure 7). However, there is still a necessity for improvement aiming to make the balance between sensitivity and specificity more sophisticated to eliminate the number of false positives; thus, the model will become more reliable and practical in clinical practice.

Figure 7. Loss vs. accuracy

Table 5. Classification report of NasNetMobile

	Precision	Recall	F1-score
0	0.65	0.95	0.77
1	0.87	0.38	0.53
Accuracy		0.69	
Sensitivity		0.38	
Specificity		0.95	
AUC-ROC		0.82	
Macro avg	0.76	0.67	0.65

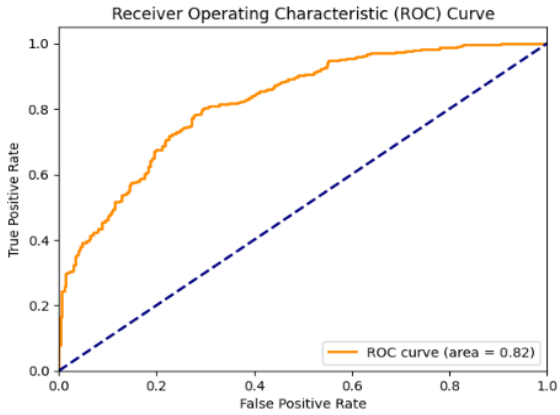

Figure 8. ROC of NasNet

5.3 NASNetMobile model performance

A number of key performance metrics were used to evaluate the diagnostic capacity available in our NASNetMobile model for skin cancer detection The achieved overall accuracy of the model was 69%, reflecting the model overall moderate level of correctness in identifying/ classifying skin lesions.

Our NASNetMobile model performance at skin cancer diagnosis is analysed, through key metrics, and this will give us a clear-cut idea regarding the effectiveness of our model. The model had an overall accuracy of 69% in making a correct identification on whether the image contained a skin lesion or not. Sensitivity of the NASNetMobile algorithm, which equals true positive rate, was 38.33%. Humane Medicine Organization writes that a sensitivity of merely above 98% is acceptable. Consequently, a sensitivity of slightly more than 38% means that the model fails to recognize a significant percent of Malignant images. This is a critical issue because correct early detection is the goal of melanoma diagnosing in clinical practice. On the other hand, specificity of the algorithm was 95.27%. This value is acceptable because the lower the number of false positives, the fewer extra treatment patients go through, and the fewer states of confusion occur. However, it is twice lower than sensitivity; hence, the algorithm's proclivity for false negatives could be concluded from these numbers. The Precision was 87% for the class malignant and 65% for the class benign. The precision of 87% seems pretty good, meaning that if the model recognizes a case as Malignant, it is correct 8 times out of 9. The F1-score calculated as 53% for malignant and 77% for benign concerning the balance between precision and sensitivity as shown in Table 5. A worse F1-score for malignant implies the model's low capability of combining correct Malignant predictions and correct Malignant predictions.

The area under the ROC curve (AUC) indicates the model's moderate discriminative ability, namely 0.8228 An AUC closer to 1.0 would be desirable, because that would reflect increasing ability of the model to discriminate between benign and malignant lesions; an AUC of 0.82 indicates there is still substantial room for improvement prev=regrTest (negative). These output/functions are presented in Figure 8 below: ROC curves are a very useful tool to represent with the help of graphic how well some classifier is capable to discriminate between classes at different threshold settings. Thus, considering the shape of curve and our AUC value being merely 0.82, the curves show that model has some discriminative capability but is far from optimal for clinical use.

From the loss curves in Figure 9, we can observe that both training and validation set is increasing quite well but a little higher in case of validation set which means there's clear sign of overfitting. If you look the training accuracy seems to getting better over time but validation accuracy gets worse which further highlights that we are having an overfitting problem. It appears that the model is only memorizing the training data and not generalizing which implies that there is a dire need of much better regularization techniques along with

more diverse training data.

The NASNetMobile has a great potential in skin cancer detection with high specificity but suffers from lack of sensitivity and generalization properties. This is further confirmed by the AUC scores that are fairly moderate and there is a decent amount of difference in the Training and Validation performances. Future work on developing the model can concentrate on overcoming overfitting and improving sensitivity and specificity with a better balance between them to enable more reliable application of the model for clinical purposes.

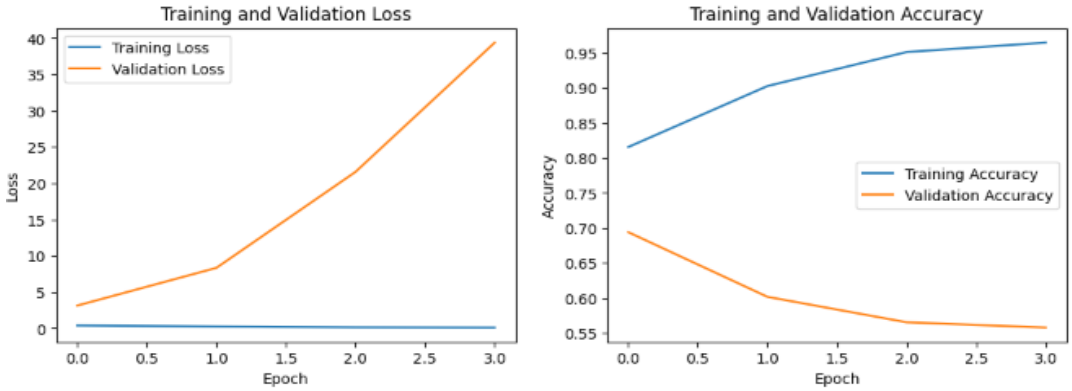

Figure 9. Loss vs. accuracy

Table 4. Classification report of VGG16

	Precision	Recall	F1-score
0	0.88	0.66	0.75
1	0.68	0.90	0.78
Accuracy		0.77	
Sensitivity		0.89	
Specificity		0.65	
AUC-ROC		0.80	
Macro avg	0.78	0.78	0.76

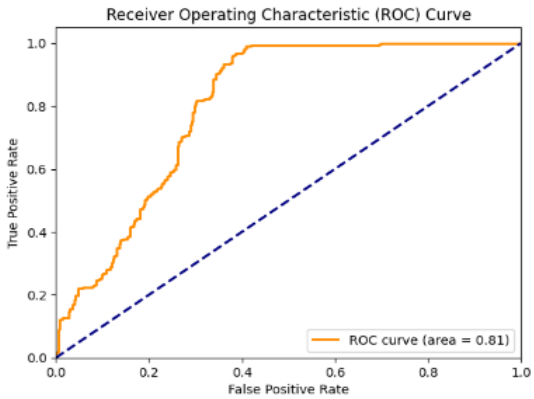

Figure 10. ROC of VGG16

5.4 VGG16

The above mentioned indexes explain in assessment of the VGG16 model used for skin cancer detection, which demonstrate its capability to enable comprehensive diagnosis. The overall accuracy for the model was 77%, which means we have a decent classification in place. Sensitivity(recall) of model, i.e. True Positive rate was 89.67%. The high sensitivity of the model ensures that most malignant lesions are identified, which is very important in order to prompt early action and treatment. The definiteness, or true negative rate was 65.56%, which can be seen as consistent with the performed pathology; after all this AI has more difficulty in saying that a lesion is benign, resulting in higher rates of false positives, but not necessarily negatives the model had a precision of 68% for malignant class and 88% for benign class. This does indeed show that the model is relatively good if it makes a prediction of malignant, however not all those predictions are correct in Table 6. The F1-score, which can be seen as a compromise between precision and recall, was 0.78 in the case of malignant cases, and 0.75 in the case of benign cases These relatively matched F1-scores indicate moderate performance of the model in detection of both serious-prognosed and harmless lesions, but it is still far from ideal. There should be some measures to take, particularly increasing specificity in case of malignancy.

As seen in Figure 10, the area under the ROC curve (AUC) was 0.8058 indicates moderators discriminative power of DT model. Mapping flow chart: Incomplete Block Symbol, an AUC value closer to 1 indicates improved performance in differently weed out between benign and malignant lesions. An AUC of 0.81 implies that the model does have some predictive power, but there is a lot of potential for improvement. ROC curve: In next ROC (Receiver Operating Characteristic) curve and it provide a graphical plot which illustrate the operational ability of a classifier model as its discrimination thresholds varied. The shape of the curve and AUC value of 0.81 indicates that the model shows some discriminative power but is still far from ideal for clinical use. Variability is also shown by the training and validation loss curves (Figure 11) as well, there are several points in which it jumps again due to increasing validation loss, indicating some level of overfit at those moments. While the training accuracy seems to increase over time (though slow), validation values fluctuates horribly, another strong sign of "we are indeed overfitting"! All of these fluctuations suggest the necessity for better regularization methods and a more varied training set to improve generalization.

The results of the study of the VGG16 model in skin cancer detection demonstrate the high sensitivity of the tool but are insufficient in relation to specificity and generalization. The described AUC, as well as the differences in performance between the training and validation models, are with the trend that justifies the need for its additional refinement. Thus, in the context of further research, it is proposed to work in the direction of mitigating the problem of overfitting and

significantly increasing specificity, which will make it possible to achieve a better balance between sensitivity and specificity and thus develop a more reliable model for clinical use.

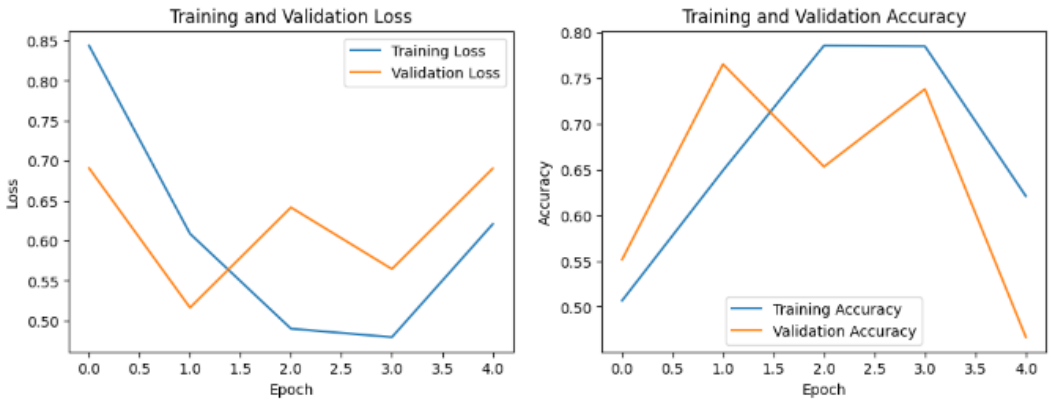

Figure 11. Loss vs. accuracy

Table 7. Classification report of Xception

	Precision	Recall	F1-score
0	0.97	0.94	0.95
1	0.93	0.96	0.95
Accuracy		0.95	
Sensitivity		0.96	
Specificity		0.93	
AUC-ROC		0.98	
Macro avg	0.95	0.95	0.95

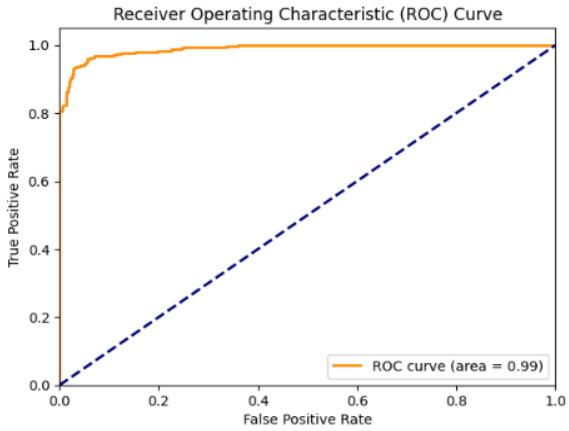

Figure 12. ROC of Xception

5.5 Xception

In this work various performance metric was examined, which gives the complete and overall idea of Xception model diagnostic result for skin cancer detection. As seen in Table 7, The model was able to achieve an overall accuracy of 95%, that means having good general accuracy in correctly identifying skin lesion. with the model having a sensitivity of 96.33% as seen above, it implies that the minority of malignant lesion will be missed making it to identify majority (most) of them and this is important for treatment also early intervention. The specificity (or true negative rate) was 93.89%, showing the level at which this model can propose that patients with benign lesions do not need to receive unnecessary treatments or feel as anxious concerning their health and well-being. The precision for the malignant class is 93%, and between the benign it goes to 97%. This follows that our model presents high correct positive rate — no matters how much false positives we have. The F1-score, which considers both precision and recall was 0.95 in malignant as well as benign cases. Both these balanced F1-scores indicate that the model is good at identifying both types of lesions — benign as well as malignant.

The area under the ROC curve (AUC) was 0.9886 (Figure 12), which also confirms excellent discriminative ability of our final model. An AUC number close to 1.0 suggests a model with high true positive rate and low false positive rate, which is the best diagnostic case you can get The ROC curve which is a plot of true positive rate vs false positive rate gives away an overall idea how well the model is able to distinguish between classes células in multiple threshold settings. With the curve shape together with AUC of 0.99, we can be sure that model is very effective in distinguishing benign from malignant lesions (as expected). Loss curves on both training and validation go down (Figure 13), indicating the learning process is going well. Accuracy curve goes up over Epochs constantly. Nonetheless, the sometimes spikes of validation loss indicate actual points in time when the model had begun to slightly overfit, which underscores the importance of regular monitoring and potential regularization methods to ensure generalization.

5.6 Comparison between models

The performances of various models developed for skin cancer detection were assessed using several key metrics including accuracy, precision, recall, F1-score, sensitivity, specificity, and area under ROC curve. The CNN model performed as the best model with the highest achievable accuracy of 97%, perfect precision for class 0 and good recall for class 1 hence high/uptake of F1-scores and a near-perfect ROC curve with 0.9962 indicating perfect ability to correctly predict skin lesions with minimal false positive and negative cases (Table 8). The Xception model performed second best with an accuracy of 95%, high precision and recall achieved for the two classes and an area under ROC curve of 0.9800 indicating good discriminative ability. The InceptionV3 model performed relatively worse than the first two models albeit with a high accuracy of 93% and an area under ROC curve of 0.9819 from which it exhibits a perfect balance of sensitivity and specificity. This was the same case for NASNetMobile model that performed worst with an accuracy of 69%, and a very low recall for class 1 hence low f1-scores and area under ROC curve of 0.8228, indicating poor analytically and

classification performance. The VGG16 model followed suit with an accuracy of 77% and area under the ROC curve of 0.8057, with the lowest sensitivity and relatively low f1-scores. Despite the model having an equal score for precision and recall the model did not perform as well as the other 3 models above. Generally, the CNN and Xception models performed better with equal accuracy, sensitivity, and specificity ratios with the other two models having areas needing improvement hence more sensitive as well as reducing false negative rates.

Figure 13. Loss vs. accuracy

Table 8. Comparative table

Model	Accuracy	Precision_0	Precision_1	Recall_0	Recall_1	F1_0	F1_1	Sensitivity	Specificity	AUC_ROC
CNN	0.97	1.00	0.94	0.94	1.00	0.97	0.97	1.0000	0.9444	0.9962
InceptionV3	0.93	0.94	0.92	0.93	0.93	0.94	0.92	0.9267	0.9333	0.9819
NASNetMobile	0.69	0.65	0.87	0.95	0.38	0.77	0.53	0.3833	0.9528	0.8228
VGG16	0.77	0.88	0.68	0.66	0.90	0.75	0.78	0.8967	0.6556	0.8057
Xception	0.95	0.97	0.93	0.95	0.90	0.95	0.95	0.9600	0.9300	0.9800

6. CONCLUSIONS

In summary, our research on SC detection has illuminated the potential of cutting-edge DL models to significantly contribute to medical diagnostics. The utilization of diverse architectures such as InceptionV3, Xception, NASNetMobile, and VGG16 has revealed the versatility and effectiveness of CNNs in accurately identifying malignant and benign skin lesions. Achieving high ACC rates, PREC, REC, and F1-Ss, these models showcase the power of leveraging artificial intelligence for medical image analysis. The comparative analysis offers a thorough comprehension of the strengths and trade-offs inherent in each model, providing valuable insights for clinicians and researchers when choosing the most appropriate approach for SC detection. This research underscores the transformative impact of ML in augmenting diagnostic capabilities and holds promise for the continued evolution of computer-aided medical diagnostics.

Looking forward, it is imperative to explore avenues that enhance the robustness and generalizability of SC detection models. One direction involves augmenting the dataset with more diverse samples, encompassing various skin types, ethnicities, and lesion presentations. Additionally, the integration of advanced data augmentation techniques and the incorporation of multi-modal data, such as dermoscopic images or patient history, could further enrich model training.

REFERENCES

[1] Deepa, R., ALMahadin, G., Sivasamy, A. (2024). Early detection of skin cancer using AI: Deciphering dermatology images for melanoma detection. AIP Advances, 14(4): 040701. https://doi.org/10.1063/5.0188187

[2] Esteva, A., Kuprel, B., Thrun, S. (2015). Deep networks for early stage skin disease and skin cancer classification. Project Report, Stanford University.

[3] Shaaban, S., Atya, H., Mohammed, H., Sameh, A., Raafat, K., Magdy, A. (2023). Skin cancer detection based on deep learning methods. In the International Conference on Artificial Intelligence and Computer Vision, Marrakesh, Morocco, pp. 58-67. https://doi.org/10.1007/978-3-031-27762-7_6

[4] Wysocki, A.B. (1999). Skin anatomy, physiology, and pathophysiology. Nursing Clinics of North America, 34(4): 777-797. https://doi.org/10.1016/S0029-6465(22)02423-9

[5] Balaha, H.M., Hassan, A.E.S. (2023). Skin cancer diagnosis based on deep transfer learning and sparrow search algorithm. Neural Computing and Applications, 35(1): 815-853. https://doi.org/10.1007/s00521-022-07762-9

[6] Ahmed, N.J. (2021). Prescribing practices of medications in the outpatient dermatology department of a public hospital. Journal of Pharmaceutical Research International, 33(1): 70-74. https://doi.org/10.9734/jpri/2021/v33i231149

[7] Skin cancers services. https://www.thegoodskinclinic.co.in/services/skin-cancers/.

[8] Binder, M., Schwarz, M., Winkler, A., Steiner, A., Kaider, A., Wolff, K., Pehamberger, H. (1995). Epiluminescence microscopy: A useful tool for the

diagnosis of pigmented skin lesions for formally trained dermatologists. Archives of Dermatology, 131(3): 286-291. https://doi.org/10.1001/archderm.1995.01690150050011

[9] Rigel, D.S., Carucci, J.A. (2000). Malignant melanoma: Prevention, early detection, and treatment in the 21st century. CA: A Cancer Journal for Clinicians, 50(4): 215-236. https://doi.org/10.3322/canjclin.50.4.215

[10] Negin, B.P., Riedel, E., Oliveria, S.A., Berwick, M., Coit, D.G., Brady, M.S. (2003). Symptoms and signs of primary melanoma: Important indicators of Breslow depth. Cancer, 98(2): 344-348. https://doi.org/10.1002/cncr.11513

[11] Mar, V., Roberts, H., Wolfe, R., English, D.R., Kelly, J.W. (2013). Nodular melanoma: A distinct clinical entity and the largest contributor to melanoma deaths in Victoria, Australia. Journal of the American Academy of Dermatology, 68(4): 568-575. https://doi.org/10.1016/j.jaad.2012.09.047

[12] Simonyan, K., Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. https://doi.org/10.48550/arXiv.1409.1556

[13] Moussa, R., Gerges, F., Salem, C., Akiki, R., Falou, O., Azar, D. (2016). Computer-aided detection of Melanoma using geometric features. In 2016 3rd Middle East Conference on Biomedical Engineering (MECBME), Beirut, Lebanon, pp. 125-128. https://doi.org/10.1109/MECBME.2016.7745423

[14] Azmi, N.F.M., Sarkan, H.M., Yahya, Y., Chuprat, S. (2016). Abcd rules segmentation on malignant tumor and benign skin lesion images. In 2016 3rd International Conference on Computer and Information Sciences (ICCOINS), Kuala Lumpur, Malaysia, pp. 66-70. https://doi.org/10.1109/ICCOINS.2016.7783190

[15] Masood, A., Al-Jumaily, A.A., Adnan, T. (2014). Development of automated diagnostic system for skin cancer: Performance analysis of neural network learning algorithms for classification. In Artificial Neural Networks and Machine Learning–ICANN 2014: 24th International Conference on Artificial Neural Networks, Hamburg, Germany, pp. 837-844. https://doi.org/10.1007/978-3-319-11179-7_105

[16] Iyatomi, H., Oka, H., Saito, M., Miyake, A., Kimoto, M., Yamagami, J., Tanaka, M. (2006). Quantitative assessment of tumour extraction from dermoscopy images and evaluation of computer-based extraction methods for an automatic melanoma diagnostic system. Melanoma Research, 16(2): 183-190. https://doi.org/10.1097/01.cmr.0000215041.76553.58

[17] Garnavi, R., Aldeen, M., Celebi, M.E., Bhuiyan, A., Dolianitis, C., Varigos, G. (2010). Automatic segmentation of dermoscopy images using histogram thresholding on optimal color channels. International Journal of Medicine and Medical Sciences, 1(2): 126-134.

[18] Garnavi, R. (2011). Computer-aided prognosis of melanoma. Doctoral dissertation, the University of Melbourne, Australia.

[19] Celebi, M.E., Kingravi, H.A., Uddin, B., Iyatomi, H., Aslandogan, Y.A., Stoecker, W.V., Moss, R.H. (2007). A methodological approach to the classification of dermoscopy images. Computerized Medical Imaging and Graphics, 31(6): 362-373. https://doi.org/10.1016/j.compmedimag.2007.01.003

[20] Ng, V.T., Fung, B.Y., Lee, T.K. (2005). Determining the asymmetry of skin lesion with fuzzy borders. Computers in Biology and Medicine, 35(2): 103-120. https://doi.org/10.1016/j.compbiomed.2003.11.004

[21] She, Z., Liu, Y., Damatoa, A. (2007). Combination of features from skin pattern and ABCD analysis for lesion classification. Skin Research and Technology, 13(1): 25-33. https://doi.org/10.1111/j.1600-0846.2007.00181.x

[22] Bono, A., Tomatis, S., Bartoli, C., Tragni, G., Radaelli, G., Maurichi, A., Marchesini, R. (1999). The ABCD system of melanoma detection: A spectrophotometric analysis of the asymmetry, border, color, and dimension. Cancer: Interdisciplinary International Journal of the American Cancer Society, 85(1): 72-77. https://doi.org/10.1002/(SICI)1097-0142(19990101)85:1<72::AID-CNCR10>3.0.CO;2-Q

[23] Maglogiannis, I., Doukas, C.N. (2009). Overview of advanced computer vision systems for skin lesions characterization. IEEE Transactions on Information Technology in Biomedicine, 13(5): 721-733. https://doi.org/10.1109/TITB.2009.2017529

[24] Shorfuzzaman, M. (2022). An explainable stacked ensemble of deep learning models for improved melanoma skin cancer detection. Multimedia Systems, 28(4): 1309-1323. https://doi.org/10.1007/s00530-021-00787-5

[25] Tembhurne, J.V., Hebbar, N., Patil, H.Y., Diwan, T. (2023). Skin cancer detection using ensemble of machine learning and deep learning techniques. Multimedia Tools and Applications, 82(18): 27501-27524. https://doi.org/10.1007/s11042-023-14697-3

[26] Imran, A., Nasir, A., Bilal, M., Sun, G., Alzahrani, A., & Almuhaimeed, A. (2022). Skin cancer detection using combined decision of deep learners. IEEE Access, 10: 118198-118212. https://doi.org/10.1109/ACCESS.2022.3220329

[27] Vineeth, J., Hemanth, S., Rao, C.V., Pavankumar, N., Jayanna, H.S., Janardhan, C.N. (2022). Skin cancer detection using deep learning. In 2022 Fourth International Conference on Cognitive Computing and Information Processing (CCIP), Bengaluru, India, pp. 1-6. https://doi.org/10.1109/CCIP57447.2022.10058685

[28] Guergueb, T., Akhloufi, M.A. (2021). Melanoma skin cancer detection using recent deep learning models. In 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Mexico, pp. 3074-3077. https://doi.org/10.1109/EMBC46164.2021.9631047

[29] Kadampur, M.A., Al Riyaee, S. (2020). Skin cancer detection: Applying a deep learning based model driven architecture in the cloud for classifying dermal cell images. Informatics in Medicine Unlocked, 18: 100282. https://doi.org/10.1016/j.imu.2019.100282

[30] Malhotra, R., Singh, P. (2023). Recent advances in deep learning models: A systematic literature review. Multimedia Tools and Applications, 82(29): 44977-45060. https://doi.org/10.1007/s11042-023-15295-z

[31] Lilhore, U.K., Simaiya, S., Sharma, Y.K., Kaswan, K.S., Rao, K.B., Rao, V.M., Alroobaea, R. (2024). A precise model for skin cancer diagnosis using hybrid U-Net and improved MobileNet-V3 with hyperparameters optimization. Scientific Reports, 14: 4299. https://doi.org/10.1038/s41598-024-54212-8

[32] Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P. (2002). SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16: 321-357.
[33] Tembhurne, J.V., Hebbar, N., Patil, H.Y., Diwan, T. (2023). Skin cancer detection using ensemble of machine learning and deep learning techniques. Multimedia Tools and Applications, 82(18): 27501-27524. https://doi.org/10.1007/s11042-023-14697-3
[34] Naqvi, M., Gilani, S.Q., Syed, T., Marques, O., Kim, H.C. (2023). Skin cancer detection using deep learning—A review. Diagnostics, 13(11): 1911. https://doi.org/10.3390/diagnostics13111911
[35] Reitermanova, Z. (2010). Data splitting. In WDS, Czech Republic, pp. 31-36.
[36] Smith, S.L., Kindermans, P.J., Ying, C., Le, Q.V. (2017). Don't decay the learning rate, increase the batch size. arXiv Preprint arXiv:1711.00489. https://doi.org/10.48550/arXiv.1711.00489
[37] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, pp. 2818-2826. https://doi.org/10.1109/CVPR.2016.308
[38] Reddy, A.S.B., Juliet, D.S. (2019). Transfer learning with ResNet-50 for malaria cell-image classification. In 2019 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, pp. 945-949. https://doi.org/10.1109/ICCSP.2019.8697909
[39] Abbas, Q., Gul, A. (2022). Detection and classification of malignant melanoma using deep features of NASNet. SN Computer Science, 4(1): 21. https://doi.org/10.1007/s42979-022-01439-9
[40] Rashid, J., Ishfaq, M., Ali, G., Saeed, M.R., Hussain, M., Alkhalifah, T., Samand, N. (2022). Skin cancer disease detection using transfer learning technique. Applied Sciences, 12(11): 5714. https://doi.org/10.3390/app12115714
[41] Manasa, K., Murthy, G.V. (2021). Skin cancer detection using VGG-16. European Journal of Molecular and Clinical Medicine, 8(1): 1419-1427.
[42] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, pp. 1251-1258. https://doi.org/10.1109/CVPR.2017.195
[43] Ogundokun, R.O., Li, A., Babatunde, R.S., Umezuruike, C., Sadiku, P.O., Abdulahi, A.T., Babatunde, A.N. (2023). Enhancing skin cancer detection and classification in dermoscopic images through concatenated MobileNetV2 and Xception models. Bioengineering, 10(8): 979. https://doi.org/10.3390/bioengineering10080979
[44] Grandini, M., Bagli, E., Visani, G. (2020). Metrics for multi-class classification: An overview. arXiv Preprint arXiv:2008.05756. https://doi.org/10.48550/arXiv.2008.05756
[45] Liang, J. (2022). Confusion matrix: Machine learning. POGIL Activity Clearinghouse, 3(4): 304.
[46] Gu, Q., Zhu, L., Cai, Z. (2009). Evaluation measures of the classification performance of imbalanced data sets. In Computational Intelligence and Intelligent Systems: 4th International Symposium, ISICA 2009, Huangshi, China, pp. 461-471. https://doi.org/10.1007/978-3-642-04962-0_53
[47] Kynkäänniemi, T., Karras, T., Laine, S., Lehtinen, J., & Aila, T. (2019). Improved precision and recall metric for assessing generative models. Advances in Neural Information Processing Systems, 32.
[48] Sokolova, M., Japkowicz, N., Szpakowicz, S. (2006). Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation. In Australasian Joint Conference on Artificial Intelligence Hobart, Australia, pp. 1015-1021. https://doi.org/10.1007/11941439_114
[49] Rehman, M., Wang, H. (2022). Comparative review of cloud computing platforms for data science workflows. arXiv Preprint arXiv:2208.14515. https://doi.org/10.48550/arXiv.2208.14515
[50] Lin, Z., Feng, L., Ardestani, E.K., Lee, J., Lundell, J., Kim, C., Owens, J.D. (2022). Building a performance model for deep learning recommendation model training on GPUS. In 2022 IEEE 29th International Conference on High Performance Computing, Data, and Analytics (HiPC), Bengaluru, India, pp. 48-58. https://doi.org/10.1109/HiPC56025.2022.00019
[51] Wegner, D., da Silveira, A.B., Marconatto, D., Mitrega, M. (2023). A systematic review of collaborative digital platforms: Structuring the domain and research agenda. Review of Managerial Science, 2023: 1-33. https://doi.org/10.1007/s11846-023-00695-0