

A Novel End-to-End Deep Learning Approach for Skin Cancer Detection Based on Web Application



Mejdal A. Alqahtani 

Department of Industrial Engineering, King Saud University, Riyadh 11362, Saudi Arabia

Corresponding Author Email: almejdal@ksu.edu.sa

Copyright: ©2024 The author. This article is published by IIETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ts.410411>

ABSTRACT

Received: 23 September 2023

Revised: 29 March 2024

Accepted: 15 April 2024

Available online: 31 August 2024

Keywords:

skin cancer, convolutional neural network (CNN), deep learning, classification, image processing, web

Skin cancer is a common, potentially fatal condition that requires early detection for successful treatment. Many cancerous cases are diagnosed in the advanced stages, which makes the chances of recovery very small, resulting in the inability to provide appropriate treatment promptly. This includes skin cancer, which causes complete damage to the affected area until it reaches the deepest layers. Previous studies have developed systems based on the diagnosis of this disease with the help of deep learning (DL), which can detect cancer in its early stages. In this study, using the Kaggle Melanoma Skin Cancer Dataset of 10000 Images dataset, which consists of over 10,000 high-quality skin lesion images, we present a novel DL method for skin cancer detection based on the DensNet121 model. Several alternative models, including DensNet121 + XGBoost Classifier, a dedicated Convolutional Neural Network (CNN) model, an ensemble model based on DensNet121, the Enhancing CNN model, and ResNet50, were also designed, implemented, and tested in addition to our main model, DensNet121. With regard to accuracy, precision, recall, F1 score, and Matthews Correlation Coefficient (MCC), our proposed model showed promising results after undergoing thorough evaluation and comparison with other recognized models. The DensNet121-based model demonstrated astounding 98% training accuracy, demonstrating its effectiveness in learning from training data. It kept up an admirable 82% validation accuracy, demonstrating its capability to handle new cases. The test's 78% accuracy rate proved that it worked well in practical situations. All three metrics—recall, precision, and F1 score—met exceptional benchmarks of 98%, demonstrating the model's prowess at identifying true positive cases and reducing false positives. Furthermore, there was 97% of the MCC indicating the high degree of accuracy between the forecast and the outcome. In the comparison analysis our model was better than ensemble model and conventional convolutional neural networks, indicating the importance of high training accuracy and generalization.

1. INTRODUCTION

Skin cancer, a consequence of DNA mutations induced by repeated exposure to ultraviolet radiation, poses a significant threat to public health [1]. Skin cancer diagnosis remains difficult despite improvements in medical imaging. The fact that a significant portion of cancer diagnoses worldwide involve skin malignancies emphasizes the importance of early detection. Melanoma, a particularly aggressive kind of skin cancer, needs special attention because it is common in young people and has the potential to spread. However, despite growing with experience, dermatologists' diagnostic accuracy still has room for improvement. According to the International Agency for Research on Cancer (IARC) report, 325,000 new cases of melanoma diagnosed in 2020 worldwide, and 57,000 people pass away due to this [2]. Melanoma has a mortality rate of 1.7 deaths per 100,000 people and an incidence rate of 3.4 cases per 100,000 people [3]. As shown in Figure 1, there is a discernible trend of more men developing melanoma than women across the majority of world regions. In some areas,

this distinction is more obvious than in others. For instance, with rates of 42 in men and 31 in women, Australia and New Zealand have the highest incidence rates per 100,000 people. Similar rates of 19 are seen in Western Europe for both sexes. The incidence rate is 18 for men and 14 for women in Northern America. Following North Europe are rates of 17 for men and 18 for women. Contrarily, melanoma continues to be relatively uncommon in many nations throughout Asia and Africa, where incidence rates typically fall below 1 per 100,000 people [4].

Melanocytes protect the deep layers of the skin from sun exposure by generating a brown pigment called melanin [5]. Frequent exposure to the sun's UV radiation, and harmful rays can induce DNA mutations, affecting the growth of skin cells and forming cancer. Therefore, cancer develops when the body's normal cells change, causing them to grow and multiply violently without control into undifferentiated cells, forming a tumor-like mass. These malignant tumors penetrate the nearby tissues and severely damage them by depriving them of nutrition and oxygen. As a result, the attacked cells either kill

or develop into cancerous cells that support the tumor and aid in its spread throughout the body.

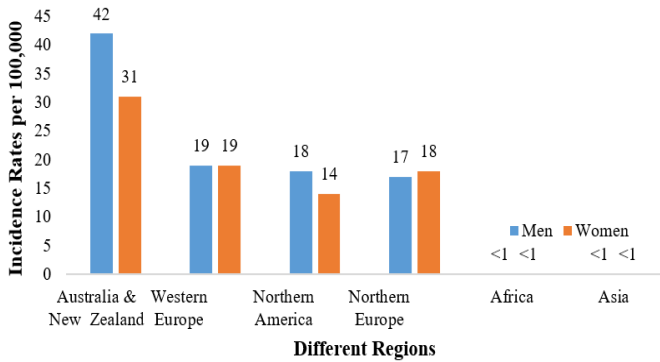


Figure 1. Melanoma incidence rates per 100,000 populations by world region and gender in 2020

Skin cancer is generally separated into two categories: melanoma and non-melanoma skin cancer, which is further separated into two more categories: basal cell and squamous cell carcinoma. Although each subtype mentioned earlier has unique traits, the present study will concentrate on melanoma skin cancer. Skin cancer from melanoma can be sporadic or spread to other areas throughout the body [6]. These malignancies are prevalent in young individuals [7]. They may be fatal if prompt treatment is not given. According to the World Health Organization, one-third of all cancer diagnoses worldwide are skin malignancies [8]. Every year, the United States records about 4.5 million cases of skin cancer, where nearly three-quarters of all skin cancer-related deaths—which total more than 10,000 every year in the United States alone—are caused by melanomas [9].

Dermatologists struggle to diagnose skin cancer accurately because dermatoscopic images can reveal many forms. Due to the similarity of many skin cancers' outward appearances, they are only exposed to a portion of all skin cancers' potential lifetime manifestations during their studies and clinical practice. Dermatologists can diagnose melanoma with an average accuracy of 62% to 80% [10, 11], but the diagnostic accuracy for those with three to five years of experience is reported to be 62%. The accuracy rate can reach up to 80% for dermatologists with more than ten years of experience. Dermatologists with less experience performed worse.⁷ Additionally, dermoscopy may be less accurate in detecting melanoma when performed by dermatologists who lack experience [10, 12, 13].

Recently, numerous past researchers utilized artificial intelligence (AI) methods to diagnose cancer precisely [14-19]. Systems called computer-aided detection (CADe) or computer-aided diagnostics (CADx) help doctors decipher and translate diagnostic medical images of patients [20]. To give the professional (radiologist) information on which to base their judgment, CAD systems process typical-looking digital images and highlight important areas that may be prospective diseases or tumors. AI, digital image processing, and radioactive image processing are all combined in the relatively new technique known as CAD. As a result, this research suggests a unique CAD system for cancer diagnosis built on a deep-learning methodology.

This research is a response to significant problems with skin cancer diagnosis. Effective treatment and the general wellbeing of patients depend on the early recognition of skin cancer, especially melanoma. However, dermatologists'

diagnoses may not always be accurate, which could lead to inaccurate or delayed assessments. Additionally, access to diagnostic expertise is frequently limited, especially in underserved areas. A user-friendly web application powered by deep learning techniques is urgently needed to address these limitations and offer accessible and reliable diagnostic support.

The major contributions of this study can be summarized as follows:

- The main contribution of this study lies in the development of a comprehensive end-to-end DL model for skin cancer diagnosis. The proposed model, DenseNet121, gives a reliable solution for classifying skin lesions into various categories like seborrheic keratosis, melanoma, and nevus.
- The current study introduces new approaches of enhancing the model. To improve the classification results, a combined technique of utilizing a neural network (DenseNet121) and a machine learning algorithm (XGBClassifier) is shown.
- In this study, the Ensemble Model-Based DensNet121 reveals an innovative strategy for using more than one model to achieve better outcomes.
- Thus, the study makes a very good observation with regard to accuracy. The proposed deep learning model surpassed the other deep learning models that are commonly employed for skin cancer recognition with a great accuracy rate of 98%.

Moreover, the study showcases the promise of data-driven innovation in healthcare, highlighting the transformative potential of DL in dermatology and cancer diagnosis. By offering a technology-driven solution, the research addresses resource constraints and provides essential diagnostic support even in areas with limited healthcare resources. In essence, the study's deep learning approach, coupled with a user-friendly web application, not only enhances skin cancer detection accuracy but also holds broader implications for healthcare accessibility and the application of artificial intelligence.

2. LITERATURE REVIEW

Deep learning methods [21-27] and traditional machine learning [28-31] are the main approaches currently used for cancer detection. Traditional machine learning techniques create cancer detection algorithms after image features are retrieved in a two-stage procedure. The authors have comprised only deep learning-based methods in the context of the work that has been presented. The feature extraction phase, which commonly uses image operators and filters to extract picture features, is where the shortcomings of traditional processes are most readily apparent. In contrast to the DL algorithm, old-fashioned feature extraction techniques do not include a learning approach; therefore, features cannot be enhanced.

Because deep learning techniques are so prevalent, some ground-breaking studies have used CNN to accurately diagnose cancer using microscopic medical images. A DL model for the sorting of cancer diagnosis was described by Pan et al. [32]. The method that is being given is an end-to-end trainable that uses CNN to carry out its categorization. A large dataset was used to test this approach, which yielded an F1 score of 94.8%. Deep learning was applied by Fakoor et al. [33] to improve the classification of cancer diagnoses. This

technique is used to identify cancer with the help of gene expression data. The overall accuracy of this approach in detecting cancer was 97.50%. Based on CNN, Rezaeilouyeh et al. [21] proposed a framework for diagnosing various cancer kinds. A phase of shearlet coefficients served as the classification's primary feature. The accuracy of the method was equal to 86%, and F1 score was equal to 89%. Danaee et al. [22] proposed a DL strategy for cancer detection by using the Stacked Denoising Autoencoder. The accuracy achieved by the authors where they used SVM as a separate classifier was 98.26%.

Among them, the comprehensive review by Dildar et al. [34] that investigate DL strategies for skin cancer diagnosis deserves a special mention. Thus, this review aims to enlighten the constantly shifting nature of this vital medical application by outlining the progress made and the challenges that still exist. Other fascinating work of interest is real-time THz imaging for skin cancer detection discussed by Lindley-Hatcher et al. [35]. Their work gives an outlook of what more can be expected out of THz imaging from having the ability to identify opportunities but also challenges.

Nahata and Singh [36] have come up with a detailed analysis of the deep learning solutions that are specifically used in the healthcare field for the detection and diagnosis of skin cancer. Their work has a big impact on how machine learning and healthcare are combined. Toğaçar et al. [37] explore the field of intelligent skin cancer detection as we shift our focus to novel methods. Their research demonstrates the enormous potential of these technologies in the field of dermatology by using autoencoders, MobileNetV2, and spiking neural networks.

For those drawn to novel approaches, Kumar et al. [38] present a fuzzy c-means clustering–inspired skin cancer detection method. This novel methodology explores novel methods that are expected to enhance the accuracy of diagnosis. Ashraf et al. [39] propose a new approach in transfer learning problem. In this situation, they provide an effective approach to leverage the capabilities of the pre-trained models using their region-of-interest based transfer learning for skin cancer detection. Al-Dmour et al. [40] present an intelligent skin cancer detection system using IoT and Fuzzy Expert System as they step into the realm of internet of things IoT. This innovative amalgamation demonstrates the possibilities of changing healthcare application with IoT technologies. Verstockt et al. [41] have also contributed in a study on the applicability of infrared thermography for skin cancer screening. This study also provides detailed and comprehensive descriptions on how the setup, process, and the tools work in this revolutionary imaging method.

Jones et al. [42] conduct a systematic review of AI and machine learning algorithms for early skin cancer detection to provide a broader perspective on AI's role in healthcare. Their research examines how AI is used in community and primary care settings, providing understanding into the game-changing potential of these tools. Nawaz et al.'s [43] research adds to the body of dermatology knowledge. Their study explores the use of DL and fuzzy k-means clustering for skin cancer recognition from dermoscopic images, illuminating the field's ongoing improvements in image-based diagnostics.

Hammad et al. [44] proposed a unique method for the detection of cancer. This method uses end-to-end deep learning, feeding the input photos straight into the deep model so that it can make the ultimate choice. The accuracy of deep CNN for cancer diagnosis is investigated in this study. The

microscopic medical images from the cancer database, which were classified as normal and abnormal, were utilized to assess our research. With respect to other deep learning models, the accuracy of the model that was given was the greatest at 99.99%."

The previous deep approaches, however, function on similar and widespread data. They also employ laborious techniques and flaws in the interpretation and application of their models in user interface-friendly applications. Therefore, our present study proposes a straightforward end-to-end DL model that reaches excellent accuracy on a short dataset, solving most of the previous issues. To advance the responsible use of artificial intelligence, it is crucial to address the current research gaps in the areas of skin cancer detection and medical applications. These gaps cover a range of crucial areas, such as dataset diversity, class imbalance reduction, the development of interpretable AI, and clinical validation via large-scale trials. The adoption of thorough evaluation metrics, ethical considerations, and seamless integration into clinical workflows are crucial areas of investigation. To ensure that everyone has access to AI tools, research should also look into multimodal approaches, longitudinal monitoring, usability, and accessibility. Investigation is necessary into the models' adaptability to various conditions as well as their robustness. Regulatory compliance and substantive patient involvement are essential for the creation of patient-centric solutions, which collectively contribute to the advancement of AI in dermatology and skin cancer detection, ultimately benefiting both patients and healthcare systems.

3. MATERIALS AND METHODS

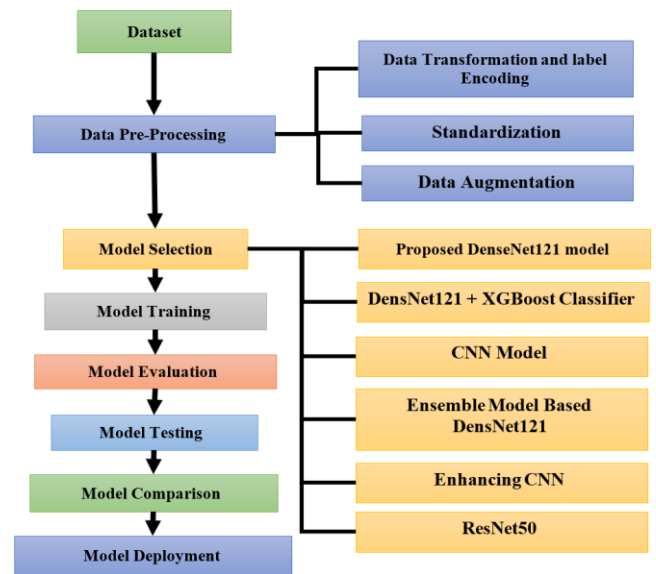


Figure 2. The proposed approach

A typical machine learning or deep learning model's structured evolution of stages is shown in Figure 2. It starts out by gathering the dataset from the Kaggle website [45]. Data preparation addresses problems like standardization and missing data. Model selection matches the selected algorithm with the objectives of the project, and model training fine-tunes the parameters to reduce error. A different dataset is used in performance evaluation to evaluate generalization. Testing demonstrates the model's accuracy in making predictions, and different models can be compared. A successful model can

then be used in practical applications. These sequential actions make sure that machine learning initiatives are approached methodically and successfully.

3.1 Dataset

The Kaggle Melanoma Skin Cancer Dataset [45] is used in this study. In order to train ML models for melanoma detection, the Kaggle dataset is a collection of skin lesion images. This dataset was developed by the International Skin Imaging Collaboration (ISIC) and was launched on Kaggle in 2017. The Melanoma Skin Cancer Dataset is a large database of dermatoscopic images of skin lesions that is made up of more than 10,000 images, where it covers both benign and melanoma. It indicates the type of lesion on the images so that supervised machine learning models can be trained from the images. Also, the images are of good quality and the dermoscopy is done by experts in the field.

3.1.1 Dataset description

The Melanoma Skin Cancer dataset is extensive and systematic which gives skin lesion images that are detailed and well arranged. The dataset is in the CSV format, and each record in the dataset corresponds to a skin lesion image. The CSV file includes distinct columns that outline crucial characteristics:

1. **Image name:** This column includes the name of the image file in order to help in correct identification and further referencing.
2. **Diagnosis:** One of the crucial elements, this column specifies the nature of the lesion, allowing the differentiation of malignant melanoma and other benign conditions.
3. **Clinical features:** Within this column resides an array of pivotal clinical features intrinsic to the lesion. These encompass pertinent aspects such as size, shape, and color, enhancing the diagnostic insights gleaned from the dataset.

4. **Dermoscopic features:** This is a critical dimension where a list of dermoscopic features related to the lesion is provided. These features include fundamental characteristics such as the existence of atypical vascular patterns or atypical pigment network, which further enrich the diagnostic approach.

All the images of the Melanoma Skin Cancer Dataset are of the same size; 256×256 pixels; which is common size of images as it does not affect the image when reduced to this size. Surprisingly, it is also visible that the mark of quality and accuracy is also reflected in the dataset because the data base contains high-resolution images of skin lesions that have been dermoscopically evaluated by experienced people. This careful selection makes the dataset to get to a level of accuracy and reliability that makes it possible to develop skin lesion analysis and diagnosis.

The proposed method was tested on 2750 images, distributed into three files: training = 2000, validation = 150, and test = 600. The images were taken from a database related to skin cancer. and they were divided into three categories of skin diseases, including seborrheic keratosis, nevus, and melanoma. An example of one of these diseases is displayed in Figure 3.

Although nevus and seborrheic keratosis are not malignant, they carry a cancer risk. There are 374 clinical dermoscopic pictures for melanoma, 1374 for nevus, and 254 for seborrheic keratosis in the train folder. Melanoma class photos total 117, nevus class images total 393, and seborrheic keratosis class images total 90 in the test folder. There are 30 photographs in the melanoma classes, 78 images in the nevus, and 42 images for seborrheic keratosis. There are 2,750 photographs in all, some of which are fuzzy and some of which have hair in them. There are a total of three important steps. Lesion segmentation comes first, then feature extraction, then lesion categorization. Sensitivity is a measure that is crucial for clinical application.

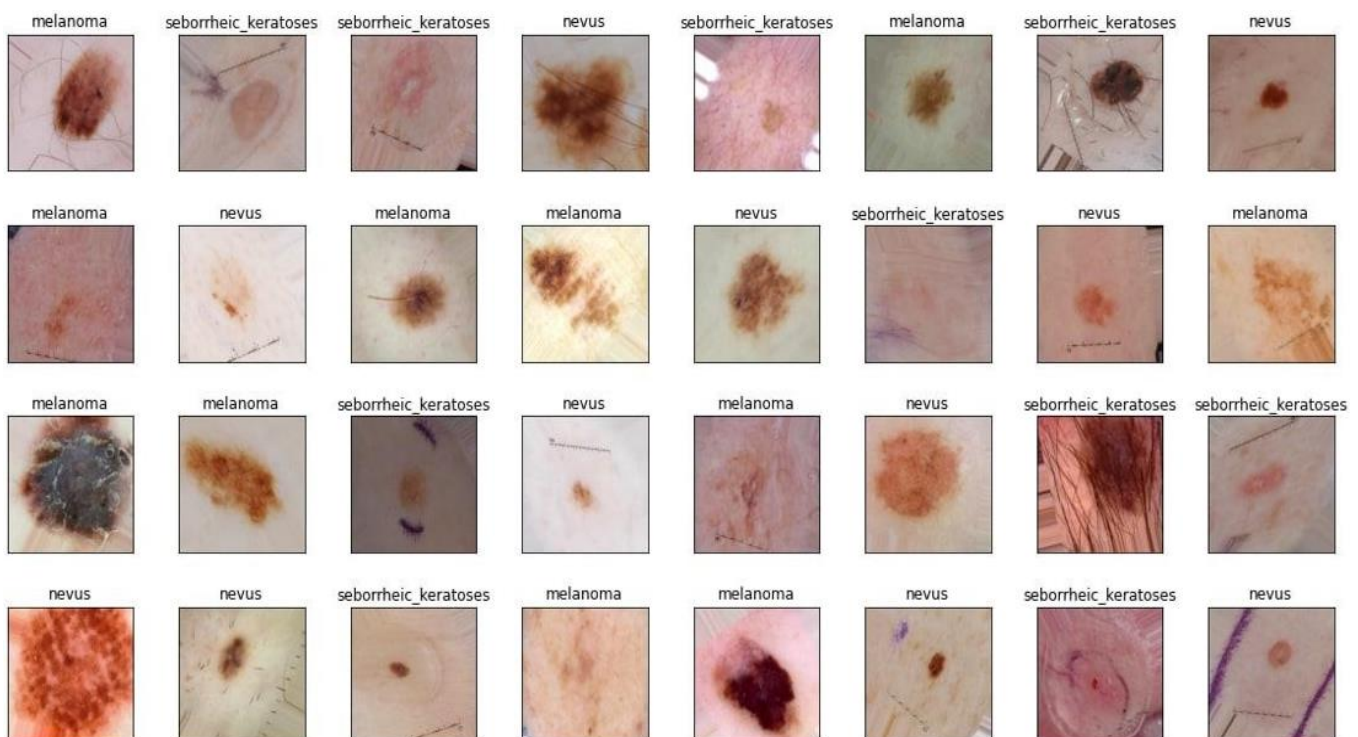


Figure 3. Example of some images from the used database

These varieties are reviewed by specialists and are among those that have been proven to impact human skin. These images are reliable evidence since they are colored and come in various sizes. Since efforts were made to get the patient's consent for their usage, concerns about privacy violations were eliminated. The data in this study had several issues, including unequal distribution of the types and diseases in each of the mentioned files. For example, in the training data, the type nevus was the most prevalent in the images, with nevus = 1372, melanoma = 374, and seborrheic keratosis = 254, respectively. This was causing rise to issues, especially when the model was fed images. The number of photos increased to 4116, which helped to address the issue considerably and enhance the outcome in the suggested model. This was done to fix the problem. The detailed description of the pre-processing of the images is presented in the subsequent section.

3.2 Dataset pre-processing

Data pre-processing is essential to modify the raw dataset in a way that is suitable for training and evaluating the model. This section just provides the way on how the dataset is prepared for the subsequent analysis. It includes various steps that guarantee the consistency, standardization, and integrity of the data.

3.2.1 Data transformation and label encoding

The main steps of the data pre-processing pipeline are the following: data transformation, and label encoding. This step involves a rigorous process of converting the categorical labels and raw skin lesion images into formats that are compatible with the next deep learning model. The first process in the system was image capturing and formatting of the raw skin lesion images which were usually obtained in several formats and resolutions. This fundamental conversion entails converting every image into a matrix of pixel values where it is possible to identify the intensity and colour of the pixel. Images are in matrices, which provides the data with a structure that is easily manageable by the model architecture.

The categorical data labels which included the possible lesion category such as melanoma, seborrheic keratosis, and nevus also underwent a label encoding process at the same time. They are in this tactical manner transformed into numerical forms that are from the said categorical labels. Seborrheic keratosis was assigned the code of "1," melanoma was assigned the code of "0,"; nevus was assigned the code of "2." In the training phase, such an encoding enables the model to understand differences between different classes of inputs and accordingly classify them. This means adopting label encoding is advantageous in several ways. It offers a very well defined and standardized format of input to the model, which in turn helps in defining a sound and systematic approach towards data interpretation. Besides, it also solves a potential integration problem by ensuring that categorical labels as inputs are converted to the numerical form of the mathematical processing in the model. Label encoding also benefits in terms of memory usage and computational performance because numbers are more easily processed.

3.2.2 Standardization of image dimensions

Another important process was to standardize the width and height of the images as it was necessary to compare the images in the dataset. This eliminated possible differences that may arise due to differences in the size of the images by

standardizing them to a size of (224, 224, 3). The smooth integration into the ensuing model architecture which results from this harmonization ensures that reliable and accurate analysis is attained.

3.2.3 Data augmentation

A significant improvement of the dataset's reliability and flexibility can be attributed to the incorporation of the ImageDataGenerator tool for data augmentation. To summarize, augmentation enhances the actual environment by deliberately introducing variations into the current dataset, which leads to the model's ability to predict outcomes for new cases more effectively.

The ImageDataGenerator tool is used to implement a wide range of augmentation methods, each of which has been painstakingly created to simulate plausible variations that skin lesion images might exhibit in real-world settings. The transformations included in this repertoire of augmentations include rotations, shifts, flips, zooms, and brightness adjustments. These small but significant changes add a bit of variability to the dataset, allowing the model to understand the subtleties of various viewpoints and angles that skin lesions can take. Data augmentation plays a main role in reducing the risk of overfitting, a phenomenon where a model becomes overly tuned to the details of the training data and becomes less capable of generalizing to new, unseen data. The output after performing the augmentation in dataset is presented in Figure 4.

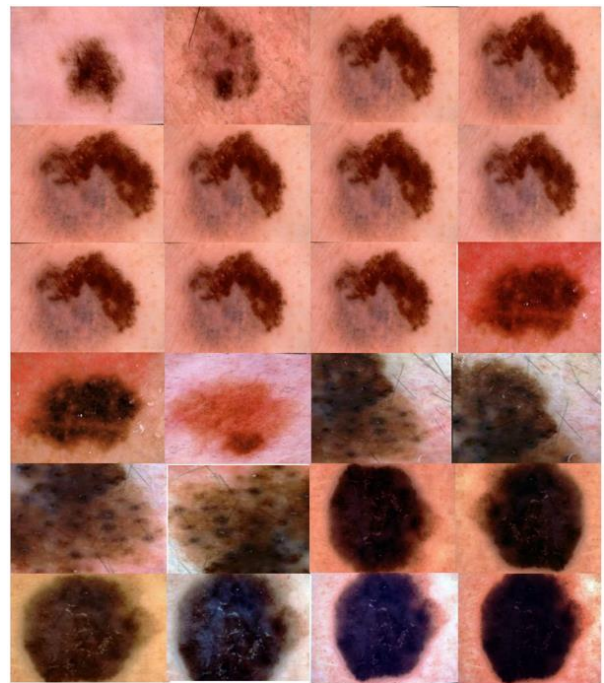


Figure 4. Output after performing data augmentation

3.2.4 Addressing class imbalances

The modality of corrections used in class imbalances within the dataset was varied in a manner that would eliminate class representation imbalance during training. This comprised of sole sampling of the minority classes, majority classes under sampling or making use of class weights during training in order to discourage the misclassification of the minority classes more heavily. To reduce the risk of biases in model predictions, careful scrutiny was done on the methods used and they were adjusted according to the results got. By being

elaborate on the training process, hyperparameter settings, model architectures and the way to handle class imbalances, your research paper will be able to provide the readers an insightful knowledge of the methodology used to design and fine tune the proposed models for skin lesion classification.

3.3 Proposed deep learning models

Our research used a comprehensive deep-learning model to identify skin cancer through images taken directly from the skin's surface. Dens121 was proposed, which was modified to suit data from input to output layers. Then, several models of neural networks were used, such as DensNet121 + XGBoost Classifier, and the CNN network was built from scratch. Ensemble Model-Based DensNet121 worked on enhancing CNN and ResNet50. Each model was tested and compared with the proposed DensNet121 model.

3.3.1 Proposed DensNet121 model

We started by building a sequential model and gradually adding layers. The variable containing the model and its batch was utilized as the first layer, and it served as a representation of the model itself. In the second layer, grouping was applied to the results of the first layer, and in a real sense, a filter with dimensions of 2×2 was created. Thus, we will have four squares, and the grouping type is max, so the four values in the filter will be replaced by the maximum value out of the four values. In the third layer, BatchNormalization was applied to the results of the second layer, which means that this layer will make the input to the next layer with a normal distribution because the layer brings the output standard deviation close to the value of 1 while the mean of the output approaches 0, which makes the input normal. In the fourth layer, a 50% sealant was applied to the results of the third layer. Then, in the fifth layer, flatten was applied to convert the outcomes of the fourth layer into a Vector so that it is easy to apply Dense since the Dense layer works at the Vector level. This vector is passed on to the next layer, the sixth layer, a type of Dense, which in turn includes 512 processing units with their biases and weights. With an activation equation of objective type, the results are passed on to the seventh layer, which is of the 50% dropout type, after which the results are finally passed to the last layer, which is the results layer and includes 3 Arithmetic units, one for each of the three data categories, with SoftMax equation as the activation equation.

Table 1 details the parameters in a proposed deep learning model that utilizes DensNet121 as its core feature extractor. The DensNet121 layer outputs a shape of (None, 7, 7, 1024) and includes 7,037,504 parameters, indicating its significant role in capturing complex patterns from the input data. This is followed by a MaxPooling2D layer that reduces the spatial dimensions to (None, 3, 3, 1024) without adding parameters. A BatchNormalization layer, keeping the same shape, introduces 4,096 parameters to stabilize and accelerate training. A Dropout layer, which helps prevent overfitting by randomly dropping units, follows with no additional parameters. The output is then flattened to a vector of 9,216 units by a Flatten layer, again without adding parameters. This vector is processed by a Dense layer, which reduces the dimension to 512 and adds 4,719,104 parameters, indicating a high-capacity fully connected layer. Another Dropout layer is used at this stage to further regularize the model. The final Dense layer outputs 3 units for classification, with an additional 1,539 parameters.

Table 1. The parameters used in the proposed model

Layer (Type)	O/P Shape	Number of Parameters
Densenet121 (Functional)	(None, 7, 7, 1024)	7037504
Max_pooling2d (MaxPooling2D)	(None, 3, 3, 1024)	0
BatchNormalization (BatchNo)	(None, 3, 3, 1024)	4096
Dense_1 (Dense)	(None, 3)	1539
Flatten (Flatten)	(None, 9216)	0
Dropout (Dropout)	(None, 3, 3, 1024)	0
Dropout_1 (Dropout)	(None, 512)	0
Dense (Dense)	(None, 512)	4719104

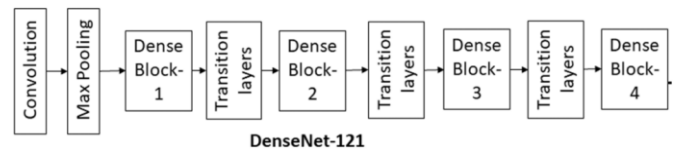


Figure 5. DenseNet121 architecture

As shown in Figure 5, Convolution sets of 3×3 and 1×1 are seen in dense blocks. Within the four dense blocks, the 3×3 and 1×1 convolutions are repeated 6, 12, 24, and 16 times, respectively. Between two thick layers lies a transition layer that is embedded. Every convolution layer within a dense block is coupled to additional convolution layers in a feed-forward fashion. 'Growth rate' is a hyperparameter used by DenseNet. Information from the prior layer is tracked by the growth rate. Between the layers, the dense connections operate via a feed-forward process. Batch normalization, one-to-one convolution, and two-to-two average pooling with a stride value of two make up the transition layer.

3.3.2 DensNet121 + XGBoost classifier

The combination of DensNet121 and the XGBoost Classifier was a multi-stage process that successfully detected skin cancer by combining the strengths of DL and machine learning paradigms. This complex strategy developed in two distinct phases, each of which contributed in its own special way to the success of the final model.

In the first step, we employed the DenseNet 121 which is a powerful pre-trained neural network. This neural network was selected because of its great ability to differentiate the features of skin images and encode the data related to cutaneous melanoma and its trends. This was done by feeding skin image data as input to DenseNet121 which served as the feature extractor. The network then captured and transmitted all these aspects to the higher level for analysis.

A smart switch was experienced when the model transitioned to the second phase; from deep learning to the XGBoost Classifier machine learning. As a result, feeding all the features extracted from the first stage into the XGBoost Classifier was easy in a structured format. This classifier, which is classified as an ensemble-based machine learning algorithm, then took control due to its high prediction capability. Its main function was to properly sort and divide the features which were extracted, in order to assign tags which would signify the possible presence of melanoma.

We were able to manipulate and carry out a variety of complex machine learning operations thanks to this paradigm-

shifting switch from deep learning to machine learning within a single model. The second stage opened up a wide range of opportunities for feature manipulation, fine-tuning, and optimization, improving the model's capacity for making decisions.

3.3.3 CNN model

The CNN is frequently used in image categorization because of its fascinating features, including automatic feature selection and end-to-end training [46]. Along with the convolution layer, CNN also heavily relies on the pooling, dropout, and dense layers. By automatically selecting features with its convolution layer, feature reduction with its pooling layer, and classification with dense layers, CNN can analyze images effectively. Figure 6 shows the CNN network's organizational structure.

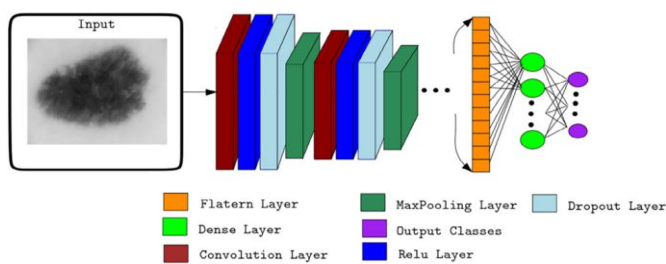


Figure 6. The CNN architecture

The relevance of each layer of CNN is different. The number of kernels and kernel size determine how many characteristics may be retrieved from each layer. The random weights that are trained during model training are used to initialize the kernel weights. The ReLU layer receives its input from the convolution layer's output. Function of nonlinear activation ReLU maintains a certain range for the convolution layer's value. Due to its simplicity, irreconcilability, and non-negativity in nature, the ReLU activation function is the most well-liked in CNN. In order to prevent the model from overfitting, the dropout layer is applied after the ReLU layer. After that, the pooling layer is used to do a down sampling on the feature map. It would be useful to make the representation resistant to even minute changes related to the input. An operation that is performed on each feature map independently is what a pooling layer does in order to generate a new set of pooled features. Max pooling, which makes the most of each patch of the feature maps, is the pooling strategy that we choose to use out of the many that are available, such as average pooling, max pooling, and so on. Extraction of the image's point and edge, in addition to other minor characteristics, is made easier with its assistance. The upper level layers of the CNN are generally layers that are totally connected to one another. The output of the pooling layer is used in these thick layers in order to arrive at a classification decision.

The softmax activation function is employed in the CNN model's last layer to provide a probability distribution for multiclass classification. In the CNN model, regularization may be used to address overfitting. By incorporating a penalty into the loss function, regularization lowers overfitting. Dropout reduces interdependent learning as a means of dealing with overfitting in CNN. After a CNN's internal weights are established, back propagation is used to adjust them to the target issue.

3.3.4 Ensemble model based DensNet121

This model used three pre-trained algorithms of the type DenseNet121 that work in parallel with each other to improve the results. Each network of the type of DenseNet121 works on extracting its features and starts with different weights from the other network that works in parallel with it, after which the results are collected. Only the best results are passed on to the rest of the network for classification and obtaining the final results.

3.3.5 Enhancing CNN

This model consists of two neural networks connected excellently to serve the purpose. The first part of the network is the processing and image enhancement part. The goal of this part is to eliminate any noise in the image and lighten the image a little because, with processing and viewing the images in training and testing, we noticed a similarity. A large percentage exists between the images despite the different categories to which the images belong. Therefore, the task of this part is to try to reduce this similarity and thus increase the effectiveness of feature extraction.

In the second part, it is a neural network that works on multiple classification or multi-class classification. It is specially designed to deal with data and separate similar features that invade the network and are obtained from the first part of the network. The network ends with the last layer, and it is responsible for producing the final results.

3.3.6 ResNet50

A pre-trained neural network, ResNet50 [47], performed the three-class classification process. The ResNet50 is a runaway network; the results that come out of it are inputs to the next layer in the overall network. Residual Network is referred to as ResNet. In order to prevent accuracy from becoming saturated and rapidly deteriorating, ResNet-50 is a network that handles the problem of accuracy deterioration as network depth grows. It has 50 layers and is more durable, and it can recognize 1000 items in a single iteration. This network does not experience the vanishing gradient problem that other deep networks do, which makes the related features and classification problem challenging to optimize. ResNet-50's strength lies in the deployment of skip-net connections, where input is often applied to the results of the network's convolutional blocks [39]. ResNet-50's intricate architecture is seen in Figure 7.

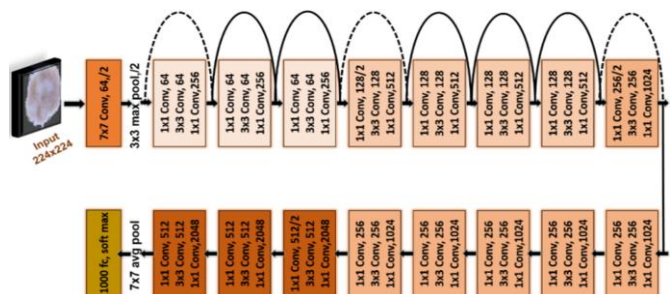


Figure 7. The ResNet-50 architecture

The ResNet-50 architecture is a type of CNN that is 50 layers deep. It's designed for image classification tasks and is known for its use of residual connections, which aid in training deeper networks by addressing the vanishing gradient issue. Here's a detailed breakdown of its structure:

Input layer: Accepts an image of size 224×224 pixels.

Initial convolution and pooling: Starts with a 7×7 convolutional layer with 64 filters and a stride of 2, followed by a 3×3 max pooling layer with a stride of 2.

Bottleneck blocks: Contains a series of bottleneck blocks, each with three layers:

- A 1×1 convolutional layer that reduces the dimension.
- A 3×3 convolutional layer that processes feature.
- Another 1×1 convolutional layer that restores the dimension.

Residual connections: Each block has a shortcut link that bypasses one or more layers. Ending Layers: Concludes with a global average pooling layer and a fully linked layer with 1000 units for the final classification.

The architecture uses batch normalization and ReLU activation after every convolutional layer. The residual connections, or shortcuts, are key to enabling the training of such deep networks by allowing gradients to flow through the network without diminishing. ResNet-50 is part of the ResNet family, which includes other variants with different depths.

3.4 K-fold cross validation

To get a better grasp of skin cancer screening methods for early detection and therapy we used a k-fold cross validation approach on the Kaggle Melanoma Skin Cancer Dataset that was based on a skin lesion images database that was thoroughly selected and comprised of more than 10 000 images.

Here, we used k-fold cross-validation as a key method for the performance of our machine learning models to ensure that. In this way, the communication cycle includes several key constituents. Firstly, we utilized an approach that was based on a smart use of loading of features and labels alongside preprocessing tasks such as handling missing values and encoding of categorical variables. As a follow up, the data was split into k fold folds, usually amounting to 5 to 10 which accommodated for the validation process. Statistically splitting the datasets to 6 folds, each one comprised of a validation set and the other folds trained our models, so we could evaluate the performance on different metrics, such as accuracy and precision. We also hyper parameter tuning in each fold to get a good performance out of the model. Outcomes produced by every fold were mixed together to represent a global score, which generally included the average performance metrics and measures of dispersion. After obtaining best responding cross-validation of the model, we have done another refinement by training them on the entire data set using optimal hyperparameters followed by performance assessment on a separate test set. By means of the targeted analysis of the cross-validation results, we reached to the understanding of the nature of our models which eventually helped us in making careful and informed decisions. Also, we do not forget the fact that in order to have robust findings we can use another random splitting algorithm.

3.5 Training process

The training of the models has been designed to maximize efficiency in terms of improving the model performance and generalization capacity. The process consisted of several critical actions, each of which was aimed towards improving reliability and efficacy of the models. At the beginning of the

process, we spent a lot of time on data preparation where the data was cleaned and divided in training, validation and test subsets. The division was done with due attention to the classes of people and equal distribution of the people within different strata. What was more, the stratified sampling methods were applied to assure the class distribution within each data set, reducing the bias chance during the training and evaluation processes. In order to add more items to the training dataset to make the model more robust, an array of data augmentation methods was used.

Which were the random rotations, translations, flips and zooms as well as light levels and contrast provided. Data augmentation not only increased the diversity of the training dataset, but it as well allowed the models to become generalists by feeding them variations in the input data, which in turn improved their learning performance. The models were implemented with Adam and SGD techniques that were considered state-of-the-art optimization algorithms at the time.

Hyperparameters like the learning rate, momentum, and weight decay were carefully tuned by accessing the validation set after many experiments and validation. Learning rate schedules like the exponential decay or step decay were used to change the original learning rate in a dynamic way during training so that the model could easily converge to the best solution. Regularization was the way to handle overfitting and increase the generalization performance. The regularization techniques were, therefore, incorporated into the training process. Dropout layers were utilized between fully connected layers to randomly peel off neurons from the network during the training phase, this in turn reduced the model's dependency on specific features and terminated the co-adaptation of neurons. Furthermore, the L2 regularization of the model's weight was done to penalize the large weight magnitudes and promote the models to become simpler.

The performance of the model is always checked on both the training set and the validation set during the training process. Performance measurements for instance, accuracy, recall, precision, and F1-score were calculated to ascertain the models' classification capability and areas for improvement. The early stopping criteria were employed according to the validation performance to ensure that overfitting was not taking place and terminating the training when it was doubtful that any further improvement would take place. While being trained, the final model is tested on the held-out test set to get a fair and unbiased estimation of the model itself. To ensure impartial assessment the test set was kept different from the training and validation sets. Model predictions on the test set were contrasted with the ground truth labels to establish and confirm the final performance metrics, thereby confirming the models' effectiveness in real-world situations. The models were systematically programmed to go through every stage of the training, which entailed data processing and augmentation, model training and testing, among others, which in turn enabled them to achieve optimal performance and robustness in image classification tasks.

4. RESULTS

We performed a thorough examination and compared our suggested DensNet121 model with other developed models to see how well it performed. Using Python and the potent libraries of Keras, Scikit-learn, and Opencv, all tests were conducted. These tests were conducted on a powerful

computing platform with an Intel Xeon CPU and a generous 128 GB of RAM. Additionally, we used a specialized NVIDIA GeForce GTX 2080 Ti GPU card's computing power to maximize the benefits of DL and speed up the training process, assuring effective and high-performance execution of the tests.

In this work, machine learning models for melanoma detection were trained using the Kaggle Melanoma Skin Cancer Dataset. The ISIC assembled this collection of skin lesion images, which was then made available on Kaggle in 2017. It contains over 10,000 high-quality photos that have been painstakingly annotated for supervised machine learning, covering both benign and malignant instances.

Table 2 covered an extensive description of the experimental setup and hardware details employed during the study. Deep learning frameworks used for these are TensorFlow version v2.15.0.post1 and Keras version 2.16, chosen as the primary ones for model development and training. A powerful NVIDIA GeForce RTX 2080 Ti GPU was utilized in this experiment, featuring a memory of 11 GB of GDDR6, 4352 CUDA cores, and 544 Tensor cores to deliver fast computation of complex tasks undergone in deep learning. In terms of CPU and memory specifications, the Intel Xeon Gold 6248R Processor with 24 cores and 48 threads running at a 3.00 GHz base clock and a max turbo frequency of 4.00 GHz, combined with 128 GB of DDR4 RAM clocked at 2933 MHz, allowed for efficient training of deep I performed these experiments on Ubuntu 20.04 LTS operating system, and installed other tools like Python (version 3.11.2), NumPy (version 1.26.0), Pandas (version 2.2.1), and Matplotlib (version 3.9.0) for dealing with data, analysis, and visualization. With such hardware and software setup, the experimental environment is more robust and efficient, thereby creating the optimal working environment for the successful execution of deep learning experiments and outcomes.

Table 2. Experimental setup and hardware details

Component	Specification
Deep Learning Frameworks	
TensorFlow	Version v2.15.0.post1
Keras	Version 2.16
GPU Specifications	
GPU Model	NVIDIA GeForce RTX 2080 Ti
CUDA Cores	4352
GPU Memory	11 GB GDDR6
Tensor Cores	544
CPU and Memory	
CPU Model	Intel Xeon Gold 6248R Processor
CPU Cores/Threads	24 Cores / 48 Threads
Memory	128 GB DDR4 RAM
Base Clock	3.00 GHz
Max Turbo Frequency	4.00 GHz
Memory Speed	2933 MHz
Operating System	
OS	Ubuntu 20.04 LTS
Additional Software	
Python	Version 3.11.2
Matplotlib	Version 3.9.0
NumPy	Version 1.26.0
Pandas	Version 2.2.1

4.1 Performance metrics

Performance metrics are essential instruments for evaluating the efficacy of classification models, particularly in disciplines like machine learning and diagnostics. Accuracy,

recall, precision, and F1 score are the four crucial metrics that were used in this study.

4.1.1 Accuracy

A classification model's overall correctness is gauged by accuracy. Both True Positives and True Negatives are computed as a percentage of all the instances in the data set. Eq. (1) shows and calculates it as a percentage. A high accuracy score means that the model can classify most instances correctly, but it may not be the best metric when dealing with an unbalanced dataset.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

TP = True Positives
 TN = True Negatives
 FP = False Positives
 FN = False Negatives

4.1.2 Precision

Precision is about the accuracy of conclusive predictions. The proportion of actual positives to the total amount of positive forecasts including True Positives and False Positives is then determined. Precision is used to determine the model's capacity not to generate false positive results. If the model indicates positive results in medical diagnostics, high levels of precision give high confidence in the accuracy of the model's results. Eq. (2) serves as its representation.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

4.1.3 Recall

The model's capacity to correctly identify all pertinent instances in the dataset is measured by recall, also referred to as sensitivity or True Positive Rate (TPR). The ratio of True Positives to all instances of actual positive data (True Positives Plus False Negatives) is calculated. Recall is another way of measuring the ability of a model to detect the true positive cases, which is vital in medical diagnosis not to overlook actual cases of certain diseases. Eq. (3) serves as its representation [29].

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

4.1.4 F1 score

The F1 Score measures the precision and the recall in the same proportion. It offers a single measure to evaluate the performance of a model and gives consideration both to FP and FN. According to Eq. (4) It is the harmonic mean of precision and recall. The F1 Score is useful when it is crucial to achieve a perfect recall of all the relevant cases as well as minimize false positives.

$$F1\ Score = \left(\frac{Precision \times Recall}{Precision + Recall} \right) \quad (4)$$

4.1.5 Matthews correlation coefficient (MCC)

In classification tasks, the MCC is a performance metric. False Negatives (FN), False Positives (FP), True Positives (TP), and True Negatives (TN) are all taken into account to provide a balanced measure of classification performance,

especially when working with unbalanced datasets. The following is the MCC formula:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (5)$$

The MCC has a range of -1 to 1, where 1 denotes perfect classification, 0 denotes arbitrary classification, and -1 denotes utter discrepancy between forecasts and actual results. In real-world applications, MCC values close to 1 indicate excellent classification performance, while values close to 0 suggest performance no better than random, and values close to -1 indicate classification that is completely incorrect. In situations where class imbalance is a concern or when the cost of FP and FN differs significantly, the MCC is a useful metric for evaluating binary classification models.

These performance indicators are crucial when assessing classification models, particularly for medical uses like skin cancer detection.

4.2 Performance of DensNet121 model

The results of model testing are displayed in Table 3. These outcomes show the parameters of the suggested model, demonstrating that the specificity, accuracy, and sensitivity were the highest. The confusion matrix of the suggested model on the test images is displayed in Figure 8. Here, it displays that the error rates are few. Figure 9 shows the training and validation accuracy and loss of DensNet121 model, for both the raining and validated data.

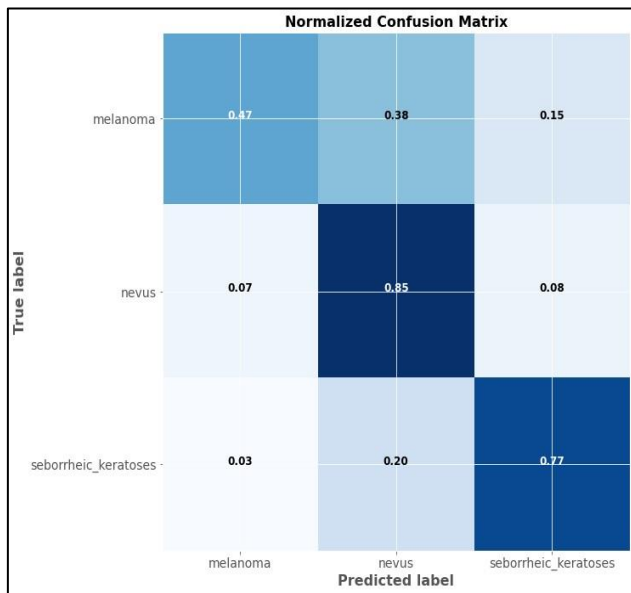
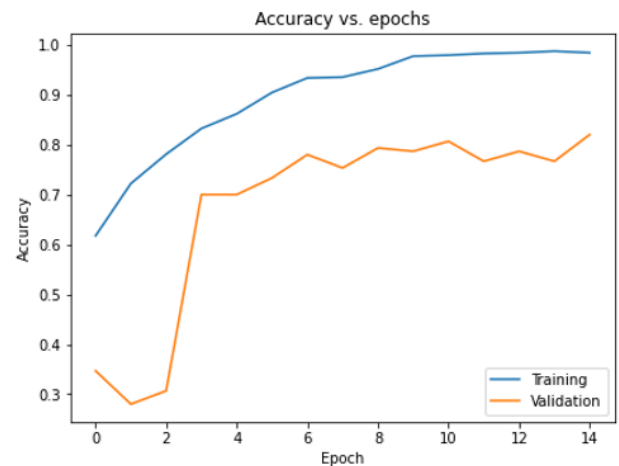


Figure 8. Confusion matrix of the proposed ensemble model

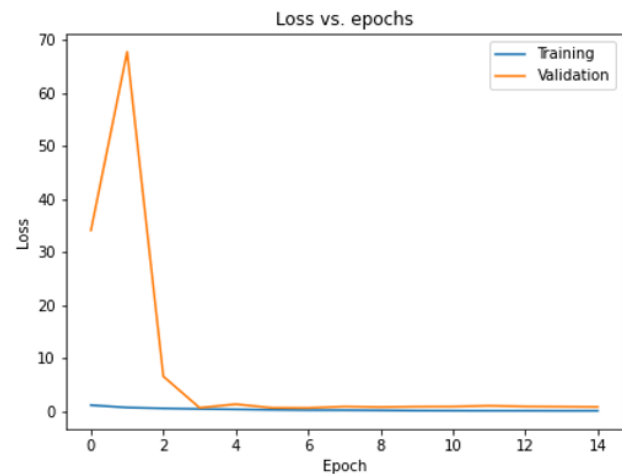
The test results for the proposed DensNet121-based DL model intended for skin cancer detection are summarized in Table 3. The performance of the model is promising across a number of important metrics. The model showed impressive accuracy during the training phase, achieving an impressive 98%, demonstrating its capacity to efficiently learn from the training data supplied. The model achieved a respectable 82% when the generalization to unseen data through validation accuracy was assessed, demonstrating its competency in handling new cases. A commendable 78% was obtained for the test accuracy, which assesses the model's applicability in

the real world, indicating that it continues to perform reasonably on new data points. The model does a good job of matching predictions with actual labels, as evidenced by the low test loss of 0.047 and slightly higher validation loss of 0.79. Notably, recall, precision, and the F1 score all have exceptional values of 98%, demonstrating the model's efficacy in capturing true positive cases and accuracy in positive forecasts. The model's strong performance is further highlighted by the MCC score of 97%, which shows a strong agreement between predictions and actual results.

The proposed DensNet121-based model performs well across a range of metrics, especially when it comes to the vital medical diagnostic metrics of precision, F1 score, recall, and MCC. The model's overall performance suggests its potential utility in aiding with the detection of skin cancer, even though test accuracy and validation loss could be improved.



(a) Accuracy of DensNet121 model



(b) Loss of DensNet121 model

Figure 9. Training and validation

Table 3. Proposed ensemble model test results

Model	DensNet121
Validation Accuracy	82%
Test Accuracy	78%
Train Accuracy	98%
Loss	0.047
Validation Loss	0.79
Recall	98%
F1 Score	98%
Precision	98%
Matthews Correlation Coefficient (MCC)	97%

4.3 Comparative analysis

The proposed DenseNet121-based model for skin cancer recognition is displayed in Table 4 and Figure 10 along with the test results of several other models, providing helpful insights into their performance. This model combines the machine learning-based XGBoost classifier with the deep learning architecture DenseNet121. Although it tested with a perfect accuracy of 100%, its test accuracy is only 70%. This decline in accuracy raises the possibility of overfitting, where the model performed admirably on the training data but had trouble extrapolating to novel, untested cases. An impressive 99.5% training accuracy was attained by a straightforward CNN model, demonstrating its capacity to learn from the training data. However, its test accuracy significantly decreased to 57%, indicating a significant difficulty in extrapolating to new data. This decrease points to the need for additional model optimization and refinement.

The training accuracy of this ensemble model, which uses several DenseNet121 instances with various weights, was 83%. However, when used with hidden data, its test accuracy of 66.0% shows a decline in performance. This suggests that although ensemble methods can be efficient, this particular configuration might call for changes. The training accuracy of the enhancing CNN layer with classification part model, which consists of two neural network parts, was 68%. Its test accuracy was 65.5%, which indicates that it maintained its performance level but found it difficult to raise it. It might require more adjusting to improve performance. The training accuracy of a pre-trained ResNet50 neural network was 68%, matching its test accuracy of 65.5%. This consistency shows that when used to analyze new data, the model consistently outperformed its training accuracy without significantly outperforming it.

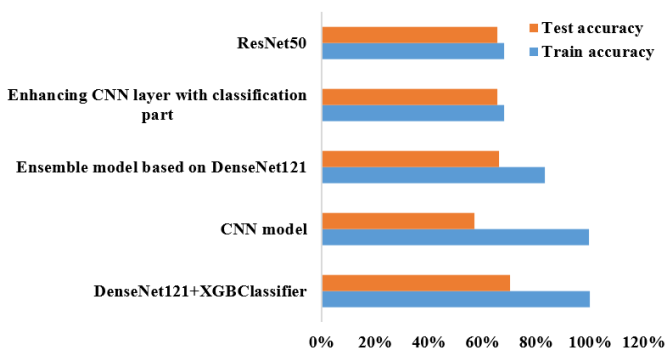


Figure 10. Train and Test results of the developed models

Table 4. Test results of the other models

Model Name	Train Accuracy	Test Accuracy
DenseNet121+XGBClassifier	100%	70%
CNN model	99.5%	57%
Ensemble model based on DenseNet121	83%	66.0%
Enhancing CNN layer with classification part	68%	65.5%
ResNet50	68%	65.5%

These findings demonstrate the various performance levels of various models. Even though some models had high training accuracy, they had trouble extrapolating to fresh, untested data, as shown by their lower test accuracy. This

demonstrates how important it is for medical diagnostics to have strong generalization in addition to high training accuracy.

The DenseNet121+XGBClassifier model employs a merge of DenseNet121 CNN architecture with a XGBoost classifier on the top. With the DenseNet121 composed of 121 layers where XGBoost got trained with parameters like maximum depth, learning rate, and number of estimators the model finally shows the complexity in its architecture and parameters. On the other hand, given this complexity, computational cost also comes into play; the depth of DenseNet121 adds to its computational cost, while XGBoost's ensemble method can be resource-heavy if used with big datasets. While the OpenCV Model requires OpenCV to be imported and integrated, the CNN Model functions as an independent standalone CNN architecture, with parameters like the number of layers, filter sizes, and learning rate considerably affecting its performance. The computational power of this model is determined by its architecture's complexity, which is often the bottleneck for utilizing large computational resources, especially in deep models.

In the Ensemble Model on DenseNet121 method, Ensemble approach is adopted, wherein multiple models of DenseNet121 may be combined. Although the ensemble model is under the same category as DenseNet121, the additional ensemble parameters is a new set of parameters introduced by the ensemble techniques. Despite the fact that the ensemble model is computationally expensive, it may actually need fewer resources in comparison with the task of training separate networks with multiple different architectures. The other model, Enhancing CNN Layer with Classification Part, brings a slight load on the computational side of things compared to training a complete CNN from the first line of code. It adds a classification layer to the basic CNN, making it a little more complex than the original. At last, the designed model utilizes the architecture of ResNet50 in its 50 layers that have skip connections. Concerning the computational requirements of the network, it is moderate since the depth is deep, but it still remains below the computational level of DenseNet121. In the end the DenseNet121+XGBClassifier, which has the highest test accuracy rate, the ensemble model based on DenseNet121, which is already very complex and resource-intensive, comes in close. While CNN and ResNet50 models are possessing relatively simpler structures and demanding less computation resource, standalone CNN and ResNet50 are the case of such models.

5. DISCUSSION

Table 4 shows the recorded results for all the models. The models employed in Table 3, which have five models and include the tile that was trained on unbalanced data, recorded a precision level of 78% when compared to the test data. We also noted that the first and most well-known model is the Densnet121+XGBCLASSIFIER since its accuracy level reached 100% in the training data but only 70% in the test data. The CNN Model, the second model, recorded an accuracy level of 99.5% in the training data while recording an accuracy level of 70% in the test data. The accuracy of the test data has risen to 57%.

The third model, entail model based on Densnet121, recorded an accuracy level of 83% in training data and 66% in test data. The "Enhancing CNN Layer with Classification

Part" model, the fourth model, achieved an accuracy level of 68% in the training data, while the accuracy level reached 65.5% in the test data. The final model, known as the RSNET50, achieved an accuracy level of 68% in the training data and an accuracy level of 65.5% in the test data. Thus, through the observation of those results recorded in Table 1 and Table 2, we note that the model known as the Densnet121 has the highest accuracy level in training and test data and was trained in balanced data.

When considering the current methods used for skin cancer detection, it is noted that the methods used are all traditional. These traditional methods are usually divided into two stages: first, the techniques are used to extract the features of the images, and then a unique algorithm is created to detect these features, as the authors are thought to have focused solely on that point. Our study differs from all the previous ones since here, we not only try to extract the features of the images but also work to classify them from one another. Ultimately, this is put in a web application to function effectively and aid in disease identification.

Jayabharathy and Vijayalakshmi [48] attempted to develop a diagnostic system for multi-category skin, and that system utilized the ResNetXt101 neural network for the MCS cancer classification, yielding results of 93.83%, 88% precision, 88% recall, and F1-score of 88. The proportions here are not very large, and more progress can be made than this. Chaturvedi et al. [49] worked on creating a mechanism for diagnosing skin cancer. Based on neural networks, deep learning was based on many models, including VGG16, VGG19, and Deep CNN. The obtained results were remarkable in terms of the network, scoring 98% accuracy for VGG16, 96% for VGG19, and 99% for the proposed DCNN.

Aburaed et al. [50] developed a method called Lesionclassifier, and the proposed system achieved 95% accuracy on the biomedical data type (ISBI) (2017) and also 93% accuracy on the ISIC (2017) data. However, this accuracy is not enough to achieve the purpose of this system. Adegun et al employed Ham10000 Dataset, CNNs, and Four Ensemble Models as the follow-up techniques. This method had an ensemble model accuracy of 92.83%. One of the strategies put up by Chaturvedi et al. [51] to deal with this issue was transfer learning using binary melanoma categorization data from the ISIC archive data and the neural network known as "VGGNET." With this method, accuracy was 81.3%, precision was 79.74%, and recall was 78.66%. In this study, we attentive on employing deep science technology, briefly touching on convolutional neural networks and the SVM algorithm. For this, we created a structure that we worked on throughout until we reached results that were not significant (85%), and the accuracy was insufficient to be used in a product.

5.1 Limitations

Although the study introduces a novel deep learning method for skin cancer detection, it has a number of drawbacks. It uses a small dataset from Kaggle that might be deficient in diversity when compared to skin conditions and populations found in the real world. Model performance may be affected by class imbalance and data quality problems, such as fuzzy or hair-contaminated images. Regarding the use of patient data, privacy concerns persist. When evaluating models, accuracy is the main metric considered, leaving out important metrics like sensitivity and specificity. The fact that generalization and

practical applications are in doubt when clinical trials and outside endorsement are lacking. It lacks real-world deployment due to complex ensemble models and interpretability problems; the feature of a web-based tool is not discussed. Despite claiming high accuracy, the study does not provide sufficient evidence of real-world generalizability or adequately address ethical issues. These limitations portray the need for further validation, beyond testing, and model transparency in the case of Medical AI, so as to determine the model's reliability, transferability, and socially responsible applicability in the clinical context.

5.2 Practical implications

The method of deep learning that was used in the study for the identification of skin cancer holds significant practical implications. For instance, in the case of melanoma, it has the potential of identifying early stages of the disease and consequently coming up with a timely treatment hence improving the patient's results. It enhances the provision of skin cancer diagnosis especially for those in the low-income bracket by developing an online diagnostic system. Moreover, the strategy of providing doctors with easily accessible tools for making accurate diagnosis contributes to reduction of the disparities in the healthcare field. It also supports medical AI research, patient self-management, and emphasizes the importance of data protection and ethical issues. Clinical integration and the requirement for additional research and validation are obstacles. In conclusion, the study's conclusions are encouraging but their successful application in clinical practice requires careful consideration of both the ethical issues and the practical difficulties.

5.3 Interpretability analysis

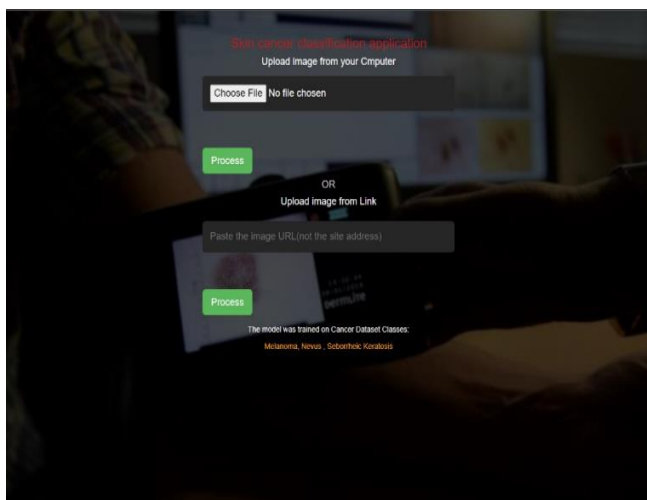
Interpretability analysis, among them methods like Grad-CAM (Gradient-weighted Class Activation Mapping), can bring about several positive outcomes related to the field of medical imaging and clinical decision-making. Interpretability analysis, by emphasizing the parts of images on which the model bases its predictions, sheds light onto the mechanisms of the model making decisions and allows the clinicians to review the decision-making process of model. The interpretability analysis usher in a number of advantages such as the improved trust and understanding of the AI-powered tools among healthcare professionals. Alongside their education and experience, clinicians usually trust their expertise and intuition during the process of diagnostic making. In this context, although AI has been a great addition in clinical practice in the recent times, the lack of clarity of AI models in their conclusions is a main problem. The interpretability analysis assists in filling this gap by supplying the clinicians with a clear graphic images reflecting those specific features that the model identifies as being related to the certain conditions. This way openness gives clinicians more confidence in the AI output and aids in more effective integration of AI in diagnostic workflows. Besides that, such analysis can serve as a guard against misinterpretation of the model's predictions or the model's failure. Through visualization of the specific image regions where the model concentrates, clinicians can judge whether the model is employing clinically pertinent attributes or not and whether or not there are some deviated patterns influencing its decisions. Such vetting can lead to the enhancement of the resilience of

AI models in medical diagnosis by referring the areas that needs to be adjusted or further training.

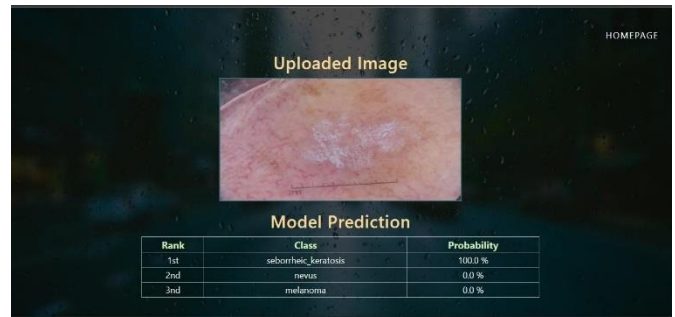
Integrating AI models, like those modeled on interpretability which is a key technique, into clinical procedures can help speed up the diagnostic processes and lead to better and more accurate diagnoses. As an example, this kind of AI-powered diagnostic tools can help clinicians by give them preliminary appraisals as well as triage the cases based on the urgency. With the ability to automatically perform routine operations, including image analysis and initial screening, AI models can relieve clinicians of taking care of the complex cases or patient care activities and focus on them. In addition to this, AI models can be incorporated into the clinical workflow flows which would help in implementing a standardized and consistent diagnostic practice across the healthcare settings. The learning process for AI algorithms can be based on a big-data approach, which involves large databases covering a variety of patient categories and clinical cases that contribute to the development of reliable diagnostics. Such harmonization can contribute to diminishing variability in diagnosing and addition to better outcomes in general. In general, interpretability analysis and medical model usage in clinical workflows can be expected to yield optimistic outcomes in the area of diagnostic processes in medicine. Via equipping doctors with AI model decision-making understanding and facilitating diagnostic workflow improvement, these approaches have the ability to enhance efficiency, accuracy, and consistency of medical diagnosis with the ultimate goal of better patient outcomes.

6. DEPLOYMENT OF THE PROPOSED MODEL

This section defines the work done to create the website that will implement the suggested model. Flask was used to create the website and also to organize and modify the website using HTML and CSS. There are two pages on the website. The first is a user request to download the affected image and post it. The website then directs you to the second page, which is the classification page, after you upload an image from your computer or attach a link to one you already have, titled 'What kind of disease did you classify your picture? Figure 11 shows the first and second pages of the website.



(a) First page



(b) Second page

Figure 11. Pages of the website

With regard to simplicity and effectiveness in terms of time, the proposed system offers a significant advantage. Patients can easily use it for the early detection of melanoma thanks to its user-friendly design, which prevents any needless complexity. This system has the potential to significantly improve patients' lives by enabling early diagnosis. There are numerous computer-aided diagnosis systems available for melanoma detection. The complexity of many of these systems, however, makes it difficult for the average person to use and navigate them effectively. These current systems frequently remain inaccessible to the general public without the direction of knowledgeable and trained individuals.

Accuracy is a defining characteristic of any successful system, and ours does not fall short in this regard. It demonstrates its ability to successfully distinguish between melanoma and non-melanoma cases with a satisfactory level of accuracy. The system's increasing accuracy creates new opportunities for useful medical applications, such as the recognition of lesions and expanded use in clinical settings. The results from our system are shown in Figure 12, which shows how well it can distinguish between melanoma and non-melanoma conditions like nevus and seborrheic keratosis. This visual illustration highlights how the system has the potential to greatly improve melanoma diagnosis and treatment, eventually helping both patients and healthcare specialists.

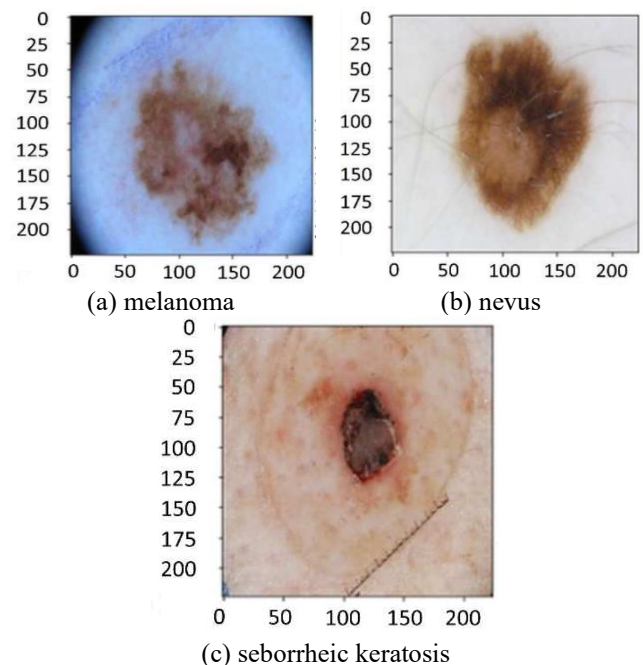


Figure 12. Accurately classified images

7. CONCLUSION

This study suggests a fresh approach to deep learning for skin cancer detection. Our study's primary objective is to develop a comprehensive method that is successful without utilizing any stages of automatic learning to achieve the highest levels. To detect skin cancer, accuracy is compared to other earlier algorithms. Seborrheic Keratosis, Nevus, and Melanoma were the three categories of the 2750 shamble pictures we worked on, and the proposed approach produced 98% resolution and 98% precision. The outcomes of our DensNet121 model's performance examination demonstrated some of its potential qualities. The model successfully learned from the supplied data during training, achieving an outstanding accuracy of 98%. Its capability to generalize to novel unseen data was evidenced by the validation accuracy of 82 percent. The model retained an accuracy of 78% in the real test scenario which showed the effectiveness of the model. The model also showed high accuracy, recall, and F1 scores of 98%, which indicates the efficiency of the model in correctly identifying positive cases. With a score of 97% the MCC significantly enhanced the performance of the model. Thus, the outcomes presented show that our proposed model based on the DensNet121 architecture can aid in the timely diagnosis of skin cancer, including melanoma. It is worth emphasizing here that there is still some room for improvement, with an eye on improving test accuracy and reducing validation loss. The effectiveness and inefficiency of the different strategies were described by comparing it with other models. Even while some models showed great training accuracy, their capacity to generalize to new data was constrained as seen by reduced test accuracy. This demonstrates the importance of good generalization and the excellent training accuracy in medical diagnostics.

Future research could further improve diagnostic precision and model generalization by investigating advanced ensemble models and addressing class imbalance issues in skin cancer detection.

ACKNOWLEDGEMENTS

The authors extend their appreciation to King Saud University for funding this work through Researchers Supporting Project number (RSP-2024R426), King Saud University, Riyadh, Saudi Arabia.

REFERENCES

- [1] Narayanan, D.L., Saladi, R.N., Fox, J.L. (2010). Ultraviolet radiation and skin cancer. *International Journal of Dermatology*, 49(9): 978-986. <https://doi.org/10.1111/j.1365-4632.2010.04474.x>
- [2] Lopes, J., Rodrigues, C.M., Gaspar, M.M., Reis, C.P. (2022). Melanoma management: from epidemiology to treatment and latest advances. *Cancers*, 14(19): 4652. <https://doi.org/10.3390/cancers14194652>
- [3] Aggarwal, P., Knabel, P., Fleischer Jr, A.B. (2021). United States burden of melanoma and non-melanoma skin cancer from 1990 to 2019. *Journal of the American Academy of Dermatology*, 85(2): 388-395. <https://doi.org/10.1016/j.jaad.2021.03.109>
- [4] Lopes, J., Rodrigues, C.M., Gaspar, M.M., Reis, C.P. (2022). Melanoma management: From epidemiology to treatment and latest advances. *Cancers*, 14(19): 4652. <https://doi.org/10.3390/cancers14194652>
- [5] American Cancer Society. <https://www.cancer.org/>, accessed on 1 January 2024.
- [6] Apalla, Z., Nashan, D., Weller, R.B., Castellsagué, X. (2017). Skin cancer: Epidemiology, disease burden, pathophysiology, diagnosis, and therapeutic approaches. *Dermatology and Therapy*, 7: 5-19. <https://doi.org/10.1007/s13555-016-0165-y>
- [7] Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 68(6): 394-424. <https://doi.org/10.3322/caac.21492>
- [8] World Health Organization. Radiation: Ultraviolet (U.V.) radiation and skin cancer. <https://www.who.int/>.
- [9] Rogers, H.W., Weinstock, M.A., Feldman, S.R., Coldiron, B.M. (2015). Incidence estimate of nonmelanoma skin cancer (keratinocyte carcinomas) in the US population, 2012. *JAMA Dermatology*, 151(10): 1081-1086. <https://doi.org/10.1001/jamadermatol.2015.1187>
- [10] Kittler, H., Pehamberger, H., Wolff, K., Binder, M. (2002). Diagnostic accuracy of dermoscopy. *The Lancet Oncology*, 3(3): 159-165. [https://doi.org/10.1016/S1470-2045\(02\)00679-4](https://doi.org/10.1016/S1470-2045(02)00679-4)
- [11] Morton, C.A., Mackie, R.M. (1998). Clinical accuracy of the diagnosis of cutaneous malignant melanoma. *British Journal of Dermatology*, 138(2): 283-287. <https://doi.org/10.1046/j.1365-2133.1998.02075.x>
- [12] 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>
- [13] Piccolo, D., Ferrari, A., Peris, K., Daidone, R., Ruggeri, B., Chimenti, S. (2002). Dermoscopic diagnosis by a trained clinician vs. a clinician with minimal dermoscopy training vs. computer-aided diagnosis of 341 pigmented skin lesions: A comparative study. *British Journal of Dermatology*, 147(3): 481-486. <https://doi.org/10.1046/j.1365-2133.2002.04978.x>
- [14] Sadoughi, F., Kazemy, Z., Hamedan, F., Owji, L., Rahmanikatiqari, M., Azadboni, T.T. (2018). Artificial intelligence methods for the diagnosis of breast cancer by image processing: A review. *Breast Cancer: Targets and Therapy*, 10: 219-230. <https://doi.org/10.2147/BCTT.S175311>
- [15] Książek, W., Hammad, M., Pławiak, P., Acharya, U.R., Tadeusiewicz, R. (2020). Development of novel ensemble model using stacking learning and evolutionary computation techniques for automated hepatocellular carcinoma detection. *Biocybernetics and Biomedical Engineering*, 40(4): 1512-1524. <https://doi.org/10.1016/j.bbe.2020.08.007>
- [16] Kumar, Y., Gupta, S., Singla, R., Hu, Y.C. (2022). A systematic review of artificial intelligence techniques in cancer prediction and diagnosis. *Archives of Computational Methods in Engineering*, 29(4): 2043-

2070. <https://doi.org/10.1007/s11831-021-09648-w>
- [17] Huang, S., Yang, J., Fong, S., Zhao, Q. (2020). Artificial intelligence in cancer diagnosis and prognosis: Opportunities and challenges. *Cancer Letters*, 471: 61-71. <https://doi.org/10.1016/j.canlet.2019.12.007>
- [18] Abdollahi, J., Keshandehghan, A., Gardaneh, M., Panahi, Y., Gardaneh, M. (2020). Accurate detection of breast cancer metastasis using a hybrid model of artificial intelligence algorithm. *Archives of Breast Cancer*, 7(1): 22-28. <https://doi.org/10.32768/abc.20207122-28>
- [19] Perincheri, S., Levi, A.W., Celli, R., et al. (2021). An independent assessment of an artificial intelligence system for prostate cancer detection shows strong diagnostic accuracy. *Modern Pathology*, 34(8): 1588-1595. <https://doi.org/10.1038/s41379-021-00794-x>
- [20] Hammad, M., Kandala, R.N., Abdelatey, A., et al. (2021). Automated detection of shockable ECG signals: A review. *Information Sciences*, 571: 580-604. <https://doi.org/10.1016/j.ins.2021.05.035>
- [21] Rezaeilouyeh, H., Mollahosseini, A., Mahoor, M.H. (2016). Microscopic medical image classification framework via deep learning and shearlet transform. *Journal of Medical Imaging*, 3(4): 044501-044501. <https://doi.org/10.1117/1.JMI.3.4.044501>
- [22] Danaee, P., Ghaeini, R., Hendrix, D.A. (2017). A deep learning approach for cancer detection and relevant gene identification. In *Pacific Symposium on Biocomputing 2017*, pp. 219-229. https://doi.org/10.1142/9789813207813_0022
- [23] Kumar, D., Jain, N., Khurana, A., Mittal, S., Satapathy, S.C., Senkerik, R., Hemanth, J.D. (2020). Automatic detection of white blood cancer from bone marrow microscopic images using convolutional neural networks. *IEEE Access*, 8: 142521-142531. <https://doi.org/10.1109/ACCESS.2020.3012292>
- [24] Khouani, A., El Habib Daho, M., Mahmoudi, S.A., Chikh, M.A., Benzineb, B. (2020). Automated recognition of white blood cells using deep learning. *Biomedical Engineering Letters*, 10: 359-367. <https://doi.org/10.1007/s13534-020-00168-3>
- [25] Yu, W., Liu, Y., Zhao, Y., et al. (2022). Deep learning-based classification of cancer cell in leptomenigeal metastasis on cytomorphologic features of cerebrospinal fluid. *Frontiers in Oncology*, 12: 821594. <https://doi.org/10.3389/fonc.2022.821594>
- [26] Hu, Z., Tang, J., Wang, Z., Zhang, K., Zhang, L., Sun, Q. (2018). Deep learning for image-based cancer detection and diagnosis— A survey. *Pattern Recognition*, 83: 134-149. <https://doi.org/10.1016/j.patcog.2018.05.014>
- [27] Lee, H., Huang, C., Yune, S., Tajmir, S.H., Kim, M., Do, S. (2019). Machine friendly machine learning: interpretation of computed tomography without image reconstruction. *Scientific Reports*, 9(1): 15540. <https://doi.org/10.1038/s41598-019-51779-5>
- [28] Dhahri, H., Al Maghayreh, E., Mahmood, A., Elkilani, W., Nagi, M.F. (2019). Automated breast cancer diagnosis based on machine learning algorithms. *Journal of Healthcare Engineering*, 2019: 4253641. <https://doi.org/10.1155/2019/4253641>
- [29] Ghasemzadeh, A., Sarbazi Azad, S., Esmaeili, E. (2019). Breast cancer detection based on Gabor-wavelet transform and machine learning methods. *International Journal of Machine Learning and Cybernetics*, 10: 1603-1612. <https://doi.org/10.1007/s13042-018-0837-2>
- [30] Karabatak, M. (2015). A new classifier for breast cancer detection based on Naïve Bayesian. *Measurement*, 72: 32-36. <https://doi.org/10.1016/j.measurement.2015.04.028>
- [31] Zheng, B., Yoon, S.W., Lam, S.S. (2014). Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms. *Expert Systems with Applications*, 41(4): 1476-1482. <https://doi.org/10.1016/j.eswa.2013.08.044>
- [32] Pan, Y., Liu, M., Xia, Y., Shen, D. (2019). Neighborhood-correction algorithm for classification of normal and malignant cells. In *ISBI 2019 C-NMC Challenge: Classification in Cancer Cell Imaging: Select Proceedings*, pp. 73-82. https://doi.org/10.1007/978-981-15-0798-4_8
- [33] Fakoor, R., Ladhak, F., Nazi, A., Huber, M. (2013). Using deep learning to enhance cancer diagnosis and classification. In *Proceedings of the 30th International Conference on Machine Learning, Atlanta, Georgia, USA*, pp. 3937-3949.
- [34] Dildar, M., Akram, S., Irfan, M., Khan, H.U., Ramzan, M., Mahmood, A.R., Alsaiari, S.A., Saeed, A.H.M., Alraddadi, M.O., Mahnashi, M.H. (2021). Skin cancer detection: A review using deep learning techniques. *International Journal of Environmental Research and Public Health*, 18(10): 5479. <https://doi.org/10.3390/ijerph18105479>
- [35] Lindley-Hatcher, H., Stantchev, R.I., Chen, X., Hernandez-Serrano, A.I., Hardwicke, J., Pickwell-MacPherson, E. (2021). Real time THz imaging - opportunities and challenges for skin cancer detection. *Applied Physics Letters*, 118(23): 230501. <https://doi.org/10.1063/5.0055259>
- [36] Nahata, H., Singh, S.P. (2020). Deep learning solutions for skin cancer detection and diagnosis. *Machine Learning with Health Care Perspective: Machine Learning and Healthcare*, 159-182. https://doi.org/10.1007/978-3-030-40850-3_8
- [37] Toğaçar, M., Cömert, Z., Ergen, B. (2021). Intelligent skin cancer detection applying autoencoder, MobileNetV2 and spiking neural networks. *Chaos, Solitons & Fractals*, 144: 110714. <https://doi.org/10.1016/j.chaos.2021.110714>
- [38] Kumar, M., Alshehri, M., AlGhamdi, R., Sharma, P., Deep, V. (2020). A DE-ANN inspired skin cancer detection approach using fuzzy c-means clustering. *Mobile Networks and Applications*, 25: 1319-1329. <https://doi.org/10.1007/s11036-020-01550-2>
- [39] Ashraf, R., Afzal, S., Rehman, A.U., et al. (2020). Region-of-interest based transfer learning assisted framework for skin cancer detection. *IEEE Access*, 8: 147858-147871. <https://doi.org/10.1109/ACCESS.2020.3014701>
- [40] Al-Dmour, N.A., Salahat, M., Nair, H.K., Kanwal, N., Saleem, M., Aziz, N. (2022). Intelligence skin cancer detection using IoT with a fuzzy expert system. In *022 International Conference on Cyber Resilience (ICCR)*, Dubai, United Arab Emirates, pp. 1-6. <https://doi.org/10.1109/ICCR56254.2022.9995733>
- [41] Verstockt, J., Verspeek, S., Thiessen, F., Tjalma, W. A., Brochez, L., Steenackers, G. (2022). Skin cancer detection using infrared thermography: Measurement setup, procedure and equipment. *Sensors*, 22(9): 3327. <https://doi.org/10.3390/s22093327>

- [42] Jones, O.T., Matin, R.N., Van der Schaar, M., et al. (2022). Artificial intelligence and machine learning algorithms for early detection of skin cancer in community and primary care settings: A systematic review. *The Lancet Digital Health*, 4(6): e466-e476. [https://doi.org/10.1016/S2589-7500\(22\)00023-1](https://doi.org/10.1016/S2589-7500(22)00023-1)
- [43] Nawaz, M., Mehmood, Z., Nazir, T., Naqvi, R.A., Rehman, A., Iqbal, M., Saba, T. (2022). Skin cancer detection from dermoscopic images using deep learning and fuzzy k-means clustering. *Microscopy Research and Technique*, 85(1): 339-351. <https://doi.org/10.1002/jemt.23908>
- [44] Hammad, M., Bakrey, M., Bakhiet, A., Tadeusiewicz, R., Abd El-Latif, A.A., Pławiak, P. (2022). A novel end-to-end deep learning approach for cancer detection based on microscopic medical images. *Biocybernetics and Biomedical Engineering*, 42(3): 737-748. <https://doi.org/10.1016/j.bbe.2022.05.009>
- [45] Melanoma Skin Cancer Dataset, Dataset Link: <https://www.kaggle.com/datasets/hasnainjaved/melano-ma-skin-cancer-dataset-of-10000-images>.
- [46] Marra, F., Gragnaniello, D., Verdoliva, L., Poggi, G. (2020). A full-image full-resolution end-to-end-trainable CNN framework for image forgery detection. *IEEE Access*, 8: 133488-133502. <https://doi.org/10.1109/ACCESS.2020.3009877>
- [47] Keras documentation: ResNet and ResNetV2. <https://keras.io/>.
- [48] Jayabharathy, K., Vijayalakshmi, K. (2022). Detection and classification of malignant melanoma and benign skin lesion using CNN. In 2022 International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN), Villupuram, India, pp. 1-4. <https://doi.org/10.1109/ICSTSN53084.2022.9761310>
- [49] Chaturvedi, S.S., Tembhurne, J.V., Diwan, T. (2020). A multi-class skin Cancer classification using deep convolutional neural networks. *Multimedia Tools and Applications*, 79(39): 28477-28498. <https://doi.org/10.1007/s11042-020-09388-2>
- [50] Aburaed, N., Panthakkan, A., Al-Saad, M., Amin, S.A., Mansoor, W. (2020). Deep convolutional neural network (DCNN) for skin cancer classification. In 020 27th IEEE International Conference on Electronics, Circuits and Systems (ICECS), Glasgow, UK, pp. 1-4. <https://doi.org/10.1109/ICECS49266.2020.9294814>
- [51] Chaturvedi, S.S., Gupta, K., Prasad, P.S. (2021). Skin lesion analyser: An efficient seven-way multi-class skin cancer classification using MobileNet. In: Hassanien, A., Bhatnagar, R., Darwish, A. (eds) *Advanced Machine Learning Technologies and Applications. AMLTA 2020. Advances in Intelligent Systems and Computing*, Springer, Singapore. https://doi.org/10.1007/978-981-15-3383-9_15