Skin Cancer Classification Using Convolution Neural Network and Meta Learning

MSc Research Project
MSc in Data Analytics

Siddharud Tevaramani
Student ID: x21156549

School of Computing
National College of Ireland

Supervisor: Noel Cosgrave

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Siddharud Tevaramani
Student ID:	x21156549
Programme:	Data Analytics
Year:	2022
Module:	MSc Research Project
Supervisor:	Noel Cosgrave
Submission Due Date:	15/12/2022
Project Title:	Skin Cancer Classification Using Convolution Neural Network and Meta Learning
Word Count:	6,232
Page Count:	19

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	26th January 2023

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Skin Cancer Classification Using Convolution Neural Network and Meta Learning

Siddharud Tevaramani
x21156549

Abstract

Skin cancer is a potentially fatal disease that can affect the entire body. There are seven types of skin cancer, including the potentially deadly melanoma. Early detection of skin cancer is crucial because delays in diagnosis can result in more serious health consequences. In the past, researchers have used machine learning algorithms to detect different types of skin cancer based on skin lesions. One such experiment involved the use of a convolutional neural network and a meta-learning technique. Highly imbalanced data is balanced by image augmentation, optimized using adam optimizerand Model has been trained using highest number of epochs by using early stopping and model checkpoint. Meta-learning is an evolving field that has shown promising output in terms of few-shot training specially when data availability is very low. In this study, multi-layer perceptron and random forest trained with a meta-learner showed an accuracy of 80%, outperforming other models that were tested.

1 Introduction

Skin cancer is a major global health concern, accounting for a significant proportion of diagnoses and treatments worldwide. One type of skin cancer, melanoma, has seen a sharp increase in incidence in the United States, with a 300% rise from 1998 to the present day. This has spurred research into the causes and potential treatments for this deadly disease(Barata and Marques; 2019). Early detection of cancer cells can greatly reduce the fatality rate for many types of cancer. If left untreated, cancer cells can spread throughout the body, leading to a higher likelihood of death. By detecting cancer at its earliest stage, it is possible to reduce the fatality rate by up to 90%, emphasizing the importance of early detection and intervention in the treatment of cancer.

The visual examination of skin lesions can be challenging due to the similarity between different types of skin cancer. In the early days of dermatological image classification, computers were used to help dermatologists gain a better understanding of skin lesions. This computer-aided diagnosis allowed clinicians to make more accurate and consistent decisions about the nature of a skin lesion. The ultimate goal of this approach is to improve the accuracy and reliability of skin cancer diagnosis, helping clinicians to make more informed treatment decisions (Höhn et al.; 2021) However, it faced two major issues. One being insufficient data and another consumption of time in the processing of the images (Alquran et al.; 2017). Later, the biopsy procedure was introduced to help dermatologists determine the stage of a skin cancer lesion. This approach involves taking a sample of tissue from the lesion and examining it under a microscope to look for cancer cells. While biopsy can provide valuable information, it is a tedious and expensive procedure that may not always be feasible. As a result, alternative methods for detecting and diagnosing skin cancer are under active investigation(Hoshyar et al.; 2011).

So it is imperative that technology that includes automatic detection and identification becomes quick and efficient solution as there is more and more requirement of specialists in dermatology, Inventing such technology would reduce or rather take out the dependencies. Further, it can be advanced to an mobile application thereby helping every household irrespective of their financial status.

The use of artificial intelligence in the early detection of skin cancer is possible. With deep convolutional neural networks and metalearner, for example, skin cancer detection systems may be developed using images of skin to identify abnormalities. Early detection of skin cancer is critical for better treatment outcomes and a longer survival time. There are several AI-based technologies in use for the diagnosis and classification of skin cancer, and this study aims to identify them. These publications will also be examined in this study for their reliability by comparing their data sets and diagnostic classifications to the performance criteria used to evaluate their models and determining how reliable they were. This will be a detailed analysis to come out with a solution of which is extensive and detailed.

Meta learning, also known as "learning to learn," is a hot topic in the field of data science research and development. It gained significant attention in 2017 and continues to evolve. (Mai et al.; 2021). This paper approach is influenced by the recent advancements in the field of meta-learning, a type of learning that is inspired by the way humans and animals process information. In meta-learning, a model learns how to improve its own learning ability by using a validation set as additional, "meta" information. This is similar to how humans and animals can adapt their learning strategies based on past experiences and knowledge.

Convolutional neural networks (ConvNets or CNNs) are commonly used for image classification tasks. In this research, the use of RGB color images and gray-level co-occurrence matrix, combined with color moment features, is proposed to improve the performance of a ConvNet for image classification. By incorporating these additional features, the proposed model may be able to more effectively classify images compared to using a ConvNet alone.

1.1 Research Question

To what extent can a meta learning algorithm combining CNN, GLCM and colour moments classify skin cancer?

1.2 Research Objectives

To provide the solution to the above research question following experiments were performed.

- Develop methodology by going through various literature
- Examine the skin lesion images, pre-processing and making it ready for modeling
- Design CNN, meantime extract GLCM features and develop meta learner.

- Comparison between the models built from CNN and Metalearner

1.3 Structure of the report

This research paper is designed section wise. Section 1 explains introduction, Section 2 walk through literature review which is again split into three parts. Section 3 talks about methodology, Section 4 covers implementation. Section 5 discusses results and evaluation and the last section 6 takes care of Conclusion and future work.

2 Related Work

In this section, various papers on the topic of skin cancer and machine learning are reviewed. Based on the findings of these papers, it is clear that there are many different approaches to solving the problem of skin cancer detection using machine learning. However, this paper has selected and implemented the model that is most suited to the given data set. The following section provides a detailed overview of the findings from the literature survey.

2.1 Convolution Neural Network for Image Classification

(Sultana et al.; 2018) discusses the reasons why convolutional neural networks (CNNs) are effective for image processing. The authors review the evolution of CNNs from LeNet-5 in 1998 to SENet in 2017. A typical CNN model consists of an input layer, followed by one or more convolution and pooling layers, and then a fully connected layer at the end. The convolution layer is the central part of the CNN, and it works by applying a kernel (i.e., a filter) to small regions of the input image to detect features that are present in those regions. The pooling layer then subsamples the convolved matrix using a method such as max pooling or mean pooling. Finally, the fully connected layer takes the output of the previous layers and combines them to produce the final output of the CNN. The authors found that SENet models are effective at identifying discriminative features in images.

As the Convolution Neural Network has been tried and tested for image processing, authors of this paper mainly rely on this technique itself. However, they tried to tweak the pre-processing thereby enhancing the performance. Initially they reduce the size of the image to 224*224 pixels. Then they augment the images by selecting among various options such as horizontal flip, zooming, rotation range, shear range, etc. Essentially every subtle change in the image shape would give different outcome. Authors have chosen rotation range of 90, shear range of 0.1, zoom range of 0.14 with horizontal flip. Three distinct steps involved are extraction of feature, feature detection and classification. Number of dropping neurons at each level has to be calculated to prevent overfitting. At the end authors, compares proposed model with VGG-16 and VGG-19, where in proposed CNN based model outperforms rest of the two models.(Rezaoana et al.; 2020)

(Jain et al.; 2021) examines the impact of quantity of training data and number of epochs and their influence on accuracy of the classification. For balancing of the data set, they augmented the images to match the each class as they found replication of images in the dataset. In addition to replicating the images only in training dataset, they modeled and compared all the six transfer learning techniques. To conclude, augmentation of the images helped in increasing the dataset to 32,0000 and thereby balancing the dataset.

Furthermore Xception Net gave good results compare to rest of the transfer learning models.

This paper mainly concentrates on bringing a process that streamlines identifying Melanoma type cancer, which is very much difficult to detect at early stages. For this, authors used Convolution Neural Network having optiimized by Whale optimization, so as to gain highest productivity. Dermquest and DermIS are used for evaluating the performance. Later on they compared the results with various models including semi-supervised models. They also considered applying this technique to openly available images from dermoscopic agencies. Sensitivity they were able to achieve was 93.5% which is significant rely on. (Zhang et al.; 2020)

Convolution Neural Networks(CNN) are used widely for hyper spectral image classification as CNNs are capable of learning features automatically, However, authors view is that CNNs only fetches spatial information and it neglects spectral data. So to tackle the issue writers propose multi-level Haar wavelet features fusion network (CNN-MHWF 2 N)(Guo et al.; 2022). 2 dimensional CNN generates spatial features and this gets combined with Haar wavelet's features to get features of both spectral and spatial.High spectral image's dimensions are reduced using factor analysis. Later 4 level decomposition are obtained using the algorithm that is based on Haar wavelet decomposition. Then, convolution layer of 4 is used to combine spatial and spectral information. Thus having higher higher information overall. They also compare the performance with other techniques and show that their method is superior.

To enrich the methodology being used in neural network design, (Huang et al.; 2020) comes up with complex networks for CNN where in they attain this by modifying the activation function. Authors implement this method mainly for magnetic resonance images of brain tumor cells. Authors start with building randomly generated graph, for this they use three algorithms Erdos-Renyi(ER), Watts-Strongatz(WS) and Barabasi-Albert (BA). Convert the randomly generated graph to directed acyclic graph based on index order method, then node operation and edge operations are performed After this output of the model. If the numbers are equal to determined module, then it forwards to classifier creator where classifier and modules are combined, followed by generating the network otherwise process repeats. With this model they were able to achieve accuracy of 95.49% also test loss of the CNN complex based networks was less when compared to ResNet, MobileNet and DenseNet.

(Sedigh et al.; 2019) presents a method for generating synthetic medical images using a Generative Adversarial Network (GAN) to improve the performance of a Convolutional Neural Network (CNN) in classifying skin cancer . The authors use the GAN to produce synthetic skin cancer images to compensate for the lack of data in the primary database used to train the CNN. The classification performance of the trained CNN without the synthetic images is near 53%, but the performance is increased to 71% when the synthetic images are added to the primary database. The authors evaluate the performance of the proposed method on the International Skin Imaging Collaboration (ISIC) dataset.

(Naeem et al.; 2020) provided a systematic literature review of recent research on melanoma classification using convolutional neural networks (CNNs). The study focuses on binary classification of melanoma and compares the accuracies of various CNN classifiers when tested on non-published datasets. The study also presents a proposed taxonomy for melanoma detection and discusses the challenges and opportunities for further research in the field.

2.2 Evolution of Meta Learning

Meta learner is a machine learning approach that learns from past machine learning approaches. Here they developed image processing recommendation algorithm using meta learning technique(Aguiar et al.; 2019). Technology include two complex steps. First major step includes modeling with meta-learning and then recommending with Meta-Learning. In the first step, Image is fed through various machine learning algorithms and then performance is evaluated, this details information becomes Meta Dataset where in Image features, Algorithm used and evaluation metrics are included. This process creates Meta-Model. Now in the 2nd step new image is fed to the system features are extracted and then Meta model kicks in, compares the its database from its learned history and then recommends best algorithms based on the image and then results into segmented image. Authors only included ML based and Gradient based algorithms. As a future work they are yet to explore on convolution neural networks.

Typical methodology starts with image processing, applying CNN method followed by Meta classifier which is 1st stage ensembler and then followed by 2nd stage orthodox ensemble techniques. Authors split 25,331 images into 95% training and 5% testing. One peculiar character that played important role is white balance processing. They were able to achieve 91% accuracy with ensemble technique compare to 87% from DenseNet121 which is a type of CNN. Authors tried their hands on cost adjustment approach and found out that it could save time and provide flexibility compared to adjusting loss function of CNN. (Lin and Lee; 2020)

(Fu et al.; 2021) uses a metalearner to improve the performance of a few-shot learning model for synthetic aperture radar (SAR) image classification. The metalearner is trained on a large dataset of SAR images and then used to train the few-shot learning model on a small dataset of SAR images. This allows the model to learn more effectively from the small dataset and improve its performance on the classification task. The authors used three transfer learning techniques to address the challenges that metalearners may encounter when learning from a small number of example images. Their approach resulted in a 1.7% improvement in performance for one-shot learning and a 2.3% improvement for five-shot learning.

(Fu et al.; 2019) presents a generalized few-shot object detection framework, called Meta-SSD, based on meta-learning. The proposed framework consists of a meta-learner and an object detector, and is able to learn general knowledge and fast adaptation strategies across many tasks. The meta-learner teaches the object detector how to learn from a few examples in just one updating step. The object detector in the framework is the Single-Shot Multibox Detector (SSD), but the framework can theoretically be used with any supervised learning detection model. The authors also construct a novel benchmark dataset for training and evaluating the proposed framework. Experiments show that Meta-SSD has promising results for few-shot object detection.

(Xue and Yu; 2020) presents a model-agnostic metalearning-based text-driven visual navigation model for unfamiliar tasks. The proposed navigation model uses meta-reinforcement learning to accumulate navigation experience from existing tasks and environments, and is able to quickly adapt to new tasks through a relatively small number of recursive trials. The model also incorporates fully convolutional instance-aware semantic segmentation and Word2vec to improve learning efficiency and accuracy. The authors evaluate the performance and generalization ability of the proposed model on the Matterport3D dataset, and show that it outperforms other traditional navigation approaches.

2.3 Background for this research

(Ali et al.; 2021) worked on categorizing benign and malignant skin lesions using deep conventional neural network (DCNN). They also evaluated the DCNN model with transfer learning models like AlexNet, VGG-16 etc. The accuracy they were able to get from DCNN model was much higher than other deep learning models and this realistically took less time. One of the challenging tasks in detecting cancer is having numerous images in different sizes, shapes and angles. Authors tackled this by doing data transformation, dimensionalality reduction, feature selection and cleaning. To overcome the over-fitting of the data, researchers tested various data augmentation methods, such as mirroring, colour shifting and mirroring. They concluded that DCNN models have higher classification rate compared to transfer learning models.

The primary objective of (Allugunti; 2022) was to demonstrate a deep learning strategy to properly diagnosing early-stage melanoma. In addition to the conventional non-parametric machine learning strategy, they used deep learning (DL) in the deep layer topologies of the convolutional neural network (CNN) algorithms. In order to evaluate the effectiveness of a CNN classifier, they used the information that could be accessed at https://dermnetnz.org/. CNN is the most widely used and successful approach of Deep Learning for the processing of images as well. For CNN models, it is feasible to build a link between the internal representation of pixel values and the two-dimensional matrix representation of those values in the output. This may be done by using the matrix representation of the values in the output. Finally, they developed Dense Conventional Network model which gave good numbers in terms of performance.

(Sharma et al.; 2022) emphasize on usage of ConvNet and cascaded ensembled deep learning models and they propose handcrafted features based multi-layer perceptron is proposed. Through the help of ensembled deep learning model authors were able to increase the accuracy by 15% when compared to orthodox CNN model. They cite that colour of the image is very crucial in identifying benign and malignant cells. There are 6 colours which are alarming with respect to malignant cells. Authors create RGB channels and calculate mean, standard deviation, skewness and kurtosis for each channels, which will generate 12 moments altogether. Methodology includes, blending of coloured images which are fed through CNN and handcrafted features such as colour moments that are fed through Multi Layer Perception to generate fully connected layer followed by softmax linear activation layer and proceeding for prediction.

3 Methodology

This study follows Knowledge Discovery in Database(KDD) for the flow of the process. The main objective of the research study is to use images from the International Skin Imaging Collaboration (ISIC) to identify and classify seven different types of skin cancers.

Figure. 1 shows the methodology diagram. It mainly highlights important steps involved. Initial step starts with Database having the skin images dataset along with meta data in the form of csv(Comma Separated Values). Later step involves extracting RGB color image which are fed to convolution neural network and meanwhile extracting Gray Level Co-occurrence Matrix. Then there are two ensemble methods to follow. One is concatenation of previous steps and another is Metalearner to provide the final output.

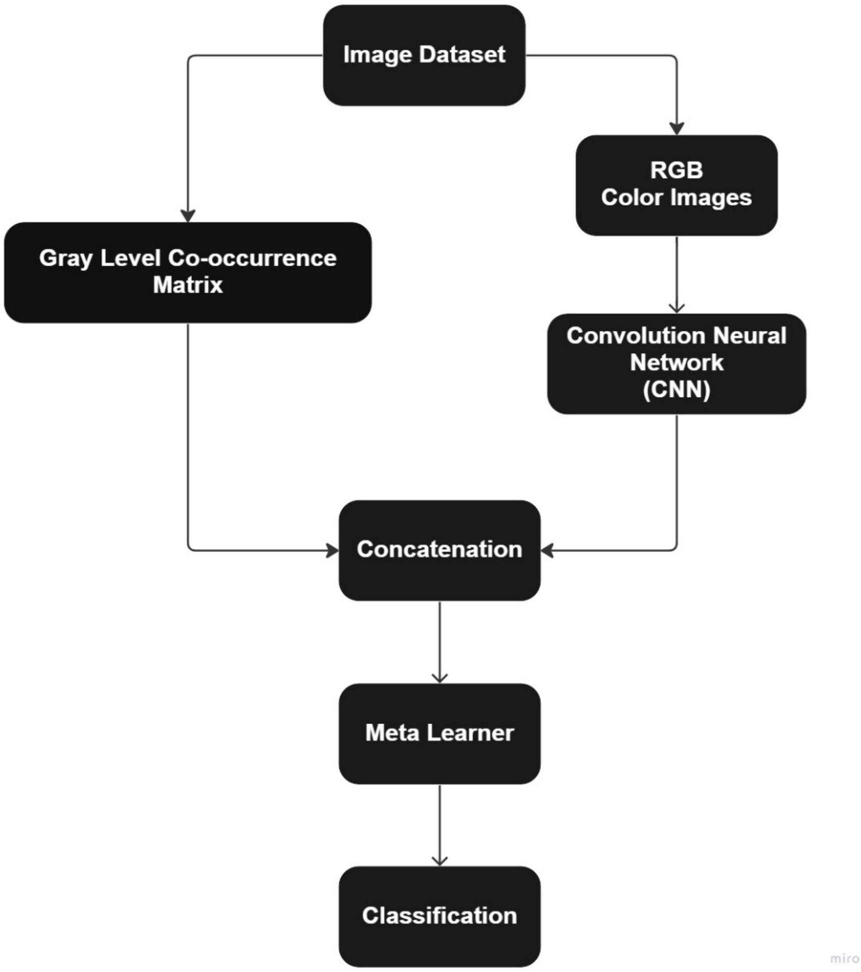


Figure 1: Architecture of the research

3.1 Data Collection

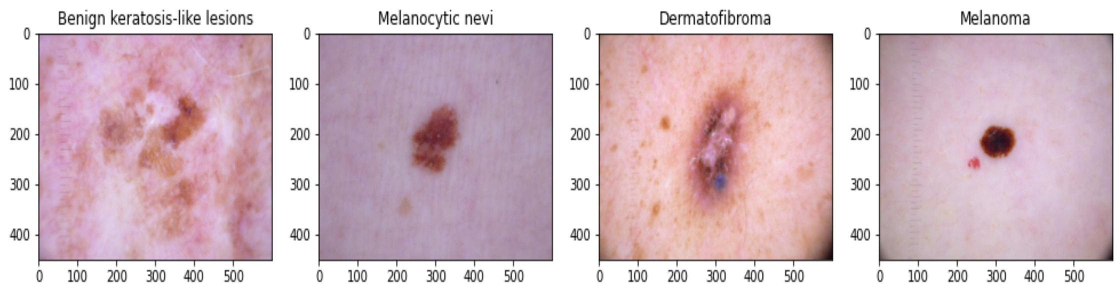
Dataset is borrowed from International Skin Imaging Collaboration(ISIC) through Harvard dataverse for the purpose study and experiments. [1]

It has 10,000 images of skin lesions split in two parts and has a Meta data that has all the details of individual image such as Age, Gender, Localization of the skin patch etc and is connected with the image with image ID. Table 3.1 shows all the variables that meta data has.

Table 3.1 Dataset Details

Attribute	**Data Type**
lesion_id	Int
image_id	Int
dx	Cat
dx_type	Cat
age	Int
sex	Cat
localization	Cat

'Lesion_id' is irrelevant for this study and 'image_id' is unique column and this image_id combines each image from 2 parts of image folder with the Meta Data csv. Next column is 'dx' which is target column and has seven categories namely 'bkl', 'nv', 'df', 'mel', 'vasc', 'bcc' and 'akiec' which stands for 'Benign keratosis-like lesions', 'Melanocytic nevi', 'Dermatofibroma', 'Melanoma', 'Vascular lesions', 'Basal cell carcinoma' and 'Actinic keratoses' respectively. Figure. 2 shows all seven types of cancer images.



The 'dx_type' variable has four categories and describes the source of cancer confirmation. Half of the cases are confirmed through histopathology, some are confirmed through follow-up examination, some through expert consensus, and the rest are confirmed by in-vivo confocal microscopy. The 'Age' column shows the patient's age at the time of cancer detection, while the 'sex' column shows the patient's gender. The 'localization' column shows where on the body the cancer was detected

[1] :https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DBW86T

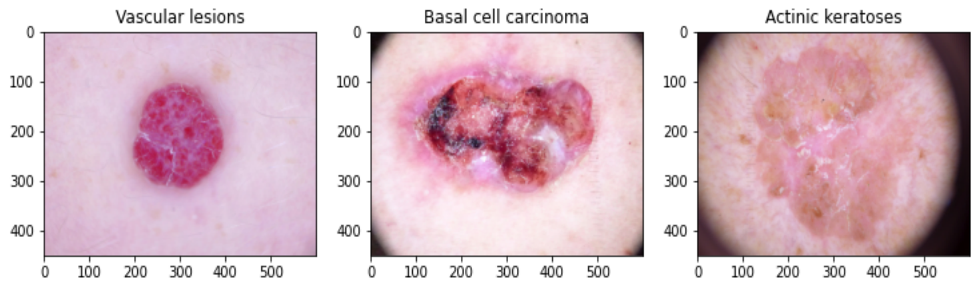


Figure 2: Images of Skin Lesion Types

3.2 Exploratory Data Analysis

Exploring the data is an important first step that helps us to better understand it. This often involves using various graphs and charts. During the analysis, researcher discovered many insights that helped in the image processing.

Figure 3 shows the age distribution of the patient at the time of cancer detection. It mostly follows normal distribution with minimal skewnewss. As shown in the histogram, affected people are in the age bracket of 40 to 50. So chances skin lesion being cancerous is more in the mid age compare to adulthood. Graph in figure 4 shows frequency of classes where Melanocytic nevi has highest occurrence making the data highly imbalanced. Balancing will be taken care in data pre-processing step.

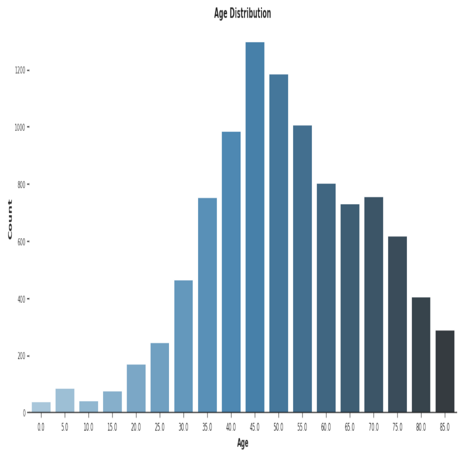


Figure 3: Age Distribution

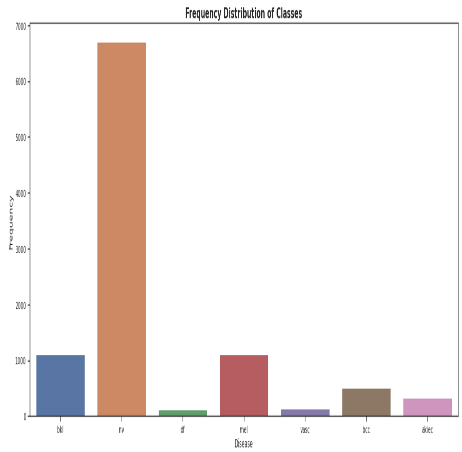


Figure 4: Frequency Distribution of Class

There is one more significant graph as shown in Figure. 5 which shows localization of skin lesion based on the gender. Males are having more cancer lesions at the back where as females have around lower extremity part which covers hip to toe. Scalp, Neck, Hand and ears have less rate.

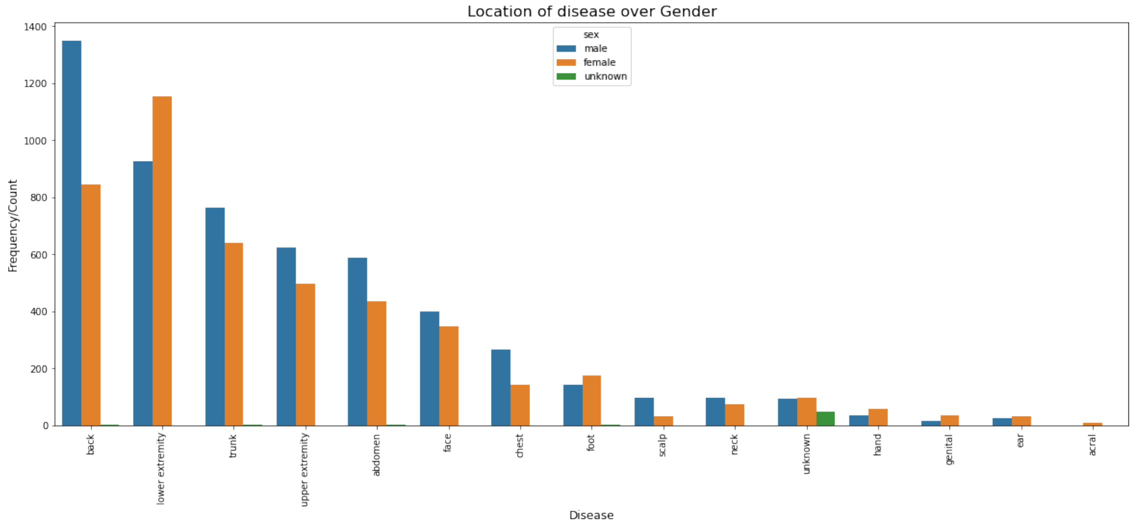


Figure 5: Localization of disease over Gender

3.3 Data Pre-Processing

In this step, Image's original resolution i.e. 600*450 is reduced to 224*224. Figure. 7 shows one of the examples. Reducing the resolution of images in deep learning helps to speed up the training process, as lower resolution images require less computational power to process. This can be especially useful when working with large datasets or when training on hardware with limited resources. Additionally, reducing the resolution of images can also help to prevent overfitting, by providing the model with less information to learn from. This can help to improve the generalization of the model, and make it more effective at making predictions on unseen data. (Dodge and Karam; 2016)

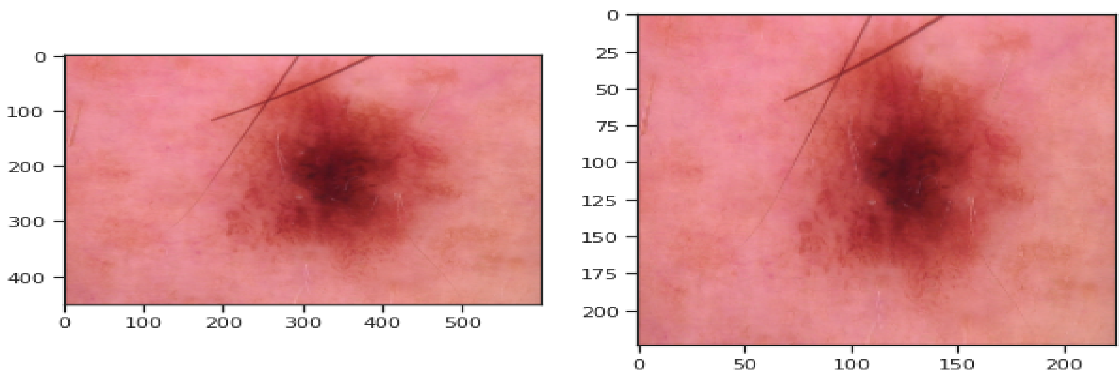


Figure 6: Reduced pixel size

Lesions have black corners which can hamper the colour dominance. There are several ways to remove black corners from an image. One approach is to use image processing techniques such as cropping or resizing to remove the black corners. Black corners are removed by reducing the pixel size and removing margin. Figure 7 shows one such example.

One major challenge in the given data is that it is highly imbalanced, with one of the classes (Melanocytic Nevi) accounting for 70% of the data. There are several ways to address this issue. One approach is to assign weights to the features in order to balance

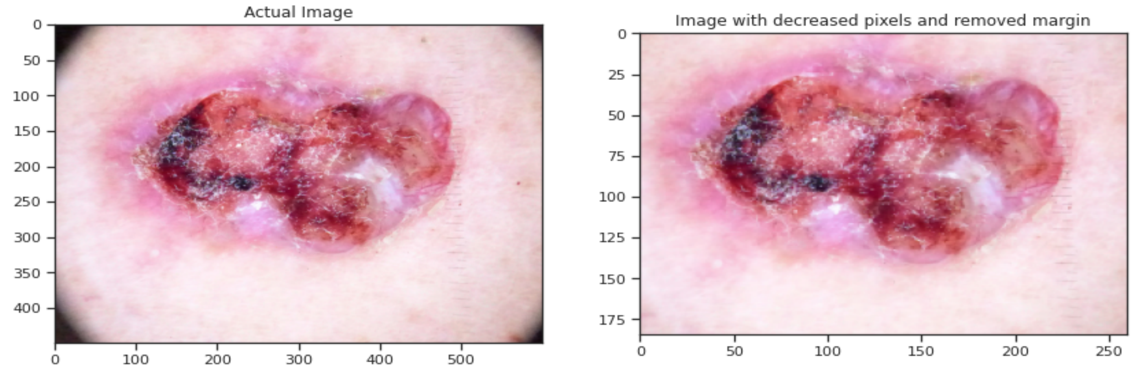


Figure 7: Black corner removal process

them. Another approach is to balance the data before starting the modeling process, for example by oversampling the minority classes or undersampling the majority class. These techniques can help improve the performance of the model and make it more robust. Researcher takes latter approach, so that issue is solved at the first place. This is done through image augmentation, where in all the classes are capped at 500. Down scaling is done through randomly selecting 500 images using resample library and same way up scaling is done by copying images randomly.

3.4 Feature Extraction

In this step, researcher check for colour dominance and extract them. In image processing, the RGB colour model is often used as a representation of the colors in an image. Each pixel in an image is represented by a combination of red, green, and blue values, which determine the color of the pixel. The RGB color model is useful for image processing because it allows for easy manipulation of individual color channels, which can be useful for a variety of tasks such as color correction, colorization, and object recognition.

Color extraction is a common technique used in image processing. It involves identifying and isolating specific colors from an image, which can be useful for tasks such as object recognition and segmentation. There are many different techniques for color extraction, such as thresholding, color space conversion, and clustering. These techniques can be used to extract a single color from an image, or to extract multiple colors and create a color palette (Ajmal et al.; 2018). Graph Figure 8 shows colour intensity graph for one of the lesions.

Meanwhile gray level co-occurrence matrix is extracted. GLCM (Gray-Level Co-occurrence Matrix) is a statistical method that is used to describe the texture of an image. It is based on the concept of co-occurrence, which refers to the occurrence of pairs of pixels with specific gray-level values in an image. GLCM features are a set of statistical measurements that are calculated from the GLCM matrix, and they are used to describe the texture of an image. Some common GLCM features include contrast, correlation, energy, and homogeneity. These features can be useful for tasks such as object recognition, classification, and segmentation. (Patel et al.; 2020)

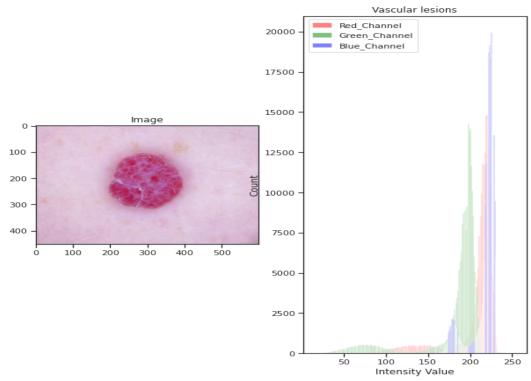


Figure 8: Colour Intensity Graph

3.5 Modeling

There are many different models that have been developed for skin cancer image classification. As seen in the literature section some of the most common approaches include convolutional neural networks (CNNs), support vector machines (SVMs), and random forests. For this study, ensemble technique will be incorporated. Initially images are classified by Convolutional Neural Network and then another layer of Metalearner is implemented.

3.6 Optimization & Evaluation

Evaluation will be done considering various graphs such as Accuracy and Loss are plotted against Epochs. Then, Hyperparameter tuning by randomization is incorporated which is a method for finding the optimal values for the hyperparameters of a machine learning model. In this approach, random values are generated for the hyperparameters, and the model is trained and evaluated using these values. The process is repeated multiple times, and the hyperparameter values that result in the best performance are selected.

4 Implementation

In this paper, the author conducted a thorough literature review and identified two main approaches to implement the ideas. The first approach is to use a well-known ConvNet model, and the second approach is to improve the performance of this model using a Metalearner. The author has discussed each of these approaches in more detail below.

4.1 Implementation with CNN

This convolutional neural network(CNN) uses the Keras library with the TensorFlow backend. It is a relatively simple model with five convolutional layers and two dense layers. The convolutional layers use a batch normalization step followed by a ReLU activation function and a max pooling layer. The dense layers use a dropout regularization step and a softmax activation function. The model is compiled using the Adam optimizer, the categorical cross-entropy loss function, and four metrics: accuracy, F1 score, precision, and recall. The model is trained using the Adam optimizer and an early stopping callback to avoid overfitting. The model is trained on image data from a directory and is validated using a validation split. The Adam optimizer is used with a learning rate of 0.00005 and default beta values. The epsilon value is set to 1e-3. These are set based on the literature that mostly The model is compiled using the Adam optimizer and the categorical cross-entropy loss function. The model is also evaluated using four metrics: accuracy, F1 score, precision, and recall.

The code then trains the model using the fit() method. The model is trained using the train generator and validated using the validation generator. The number of steps per epoch is set to the number of samples in the training set divided by the batch size. The number of validation steps is set to the number of samples in the validation set divided by the batch size. The number of epochs is set to the value of the epochs variable, and the training data is shuffled. The training process is monitored using two callbacks: early stopping and model check pointing.

4.2 Implementation of Metalearner

Meta learning is a type of machine learning that focuses on training models to learn how to learn. In other words, it is concerned with the process of learning to learn, or learning about learning. This is achieved by training the model on a variety of tasks and then using that experience to improve its ability to learn new tasks. This approach is different from traditional machine learning, where a model is trained on a specific task and then tested on that same task.

In meta learning, the model is trained on a wide range of tasks, and the performance on each task is used to improve the model's ability to learn new tasks. This allows the model to adapt quickly to new situations and learn new tasks with minimal data and computational resources. Meta learning can be seen as a form of transfer learning, where knowledge gained from one task is transferred to improve the learning of other tasks. It is also related to hierarchical learning, where a model is trained on a hierarchy of tasks, with each level of the hierarchy representing a more complex task.

Researcher has employed three classification algorithms, logistic regression, random forest and a multi layer perceptron(MLP) to transfer the knowledge gained through Meta learning, it allows the model to learn new tasks more efficiently by using the knowledge it has acquired from previous tasks. By utilizing logistic regression and a multi layer perceptron, the researcher is able to transfer the knowledge gained through Meta learning to these algorithms, allowing them to perform better on new tasks.

5 Evaluation

In this part of the project report, the performance of the models is evaluated based on their accuracy and loss in the training and test sets while model building. The report describes the results of models named ConvNet and meta learning to complete the research project. These results are briefly summarized in the report.

5.1 Convolution Neural Network

CNN model is trained with 75% data and tested on rest 25%. The metrics include the loss, accuracy, F1 score, precision, and recall. These metrics provide information about how well the model performs in terms of its ability to make corrective predictions on the test set.

The loss value is a measure of how well the model is able to predict the correct labels for the test set. A low loss value indicates that the model is making accurate predictions, while a high loss value indicates that the model struggles to make accurate predictions. In this case, the loss value is 1.1123, which suggests that the model may not be performing as desired.

The accuracy value is a measure of the overall accuracy of the model's predictions. It is calculated by dividing the number of correct predictions by the total number of predictions. In this case, the accuracy value is 0.5769, which indicates that the model is only able to make correct predictions about 57.69% of the time. This is not a particularly strong performance, and it may be worth investigating ways to improve the model's accuracy.

The F1 score, precision, and recall are metrics that are commonly used to evaluate the performance of classification model. The F1 score is a weighted average of the precision and recall, and it is a good overall measure of a model's performance. In this case, the F1 score is 0.5231, which indicates that the model is not performing very well in terms of its ability to make correct predictions. The precision and recall values are also found to be relatively low, at 0.6674 and 0.4308, respectively.

Overall, the performance metrics provided in this output suggest that the CNN model is not performing very well in terms of its ability to make accurate predictions. It may be worth investigating ways to improve the model's performance by adjusting the model architecture, using a different training dataset, or trying different hyperparameter settings.

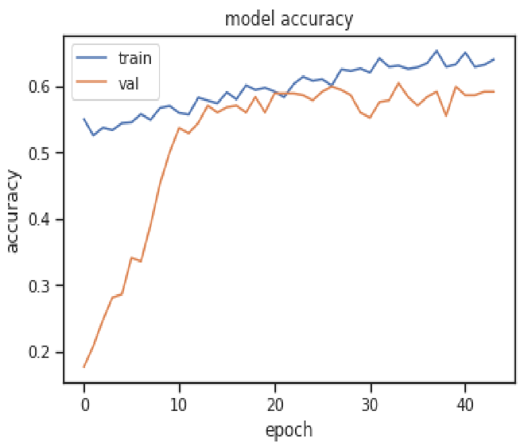


Figure 9: Model Accuracy

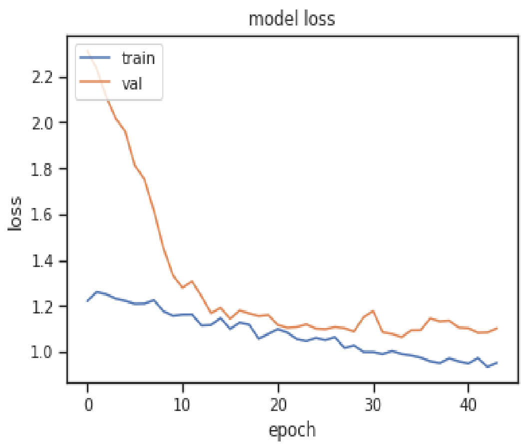


Figure 10: Model Loss

Figure. 9, Figure. 10, Figure. 11, & Figure. 12 show graphs plotted against epoch number with respect to model accuracy, model loss, model f1 score & model precision respectively.

In a accuracy and epoch graph, it is common to see the training and validation lines diverge in the early epochs and then converge again later on. This pattern occurs because the model is initially over-fitting to the training data, which leads to a high training accuracy but a low validation accuracy. As the model continues to train, it begins to generalize to the validation data and the accuracy of the model on the validation set increases.

In this case, where the training and validation lines diverge until the 15th epoch and then overlap until the last epoch of 50, this indicates that the model was initially over-

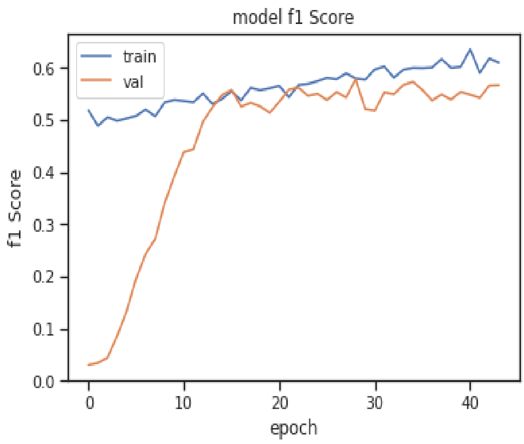


Figure 11: Model f1 score graph

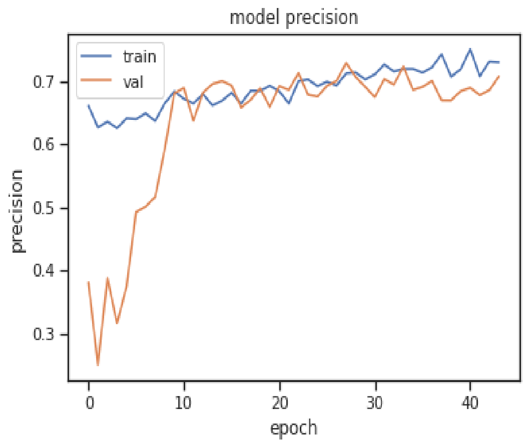


Figure 12: Model Precision Graph

fitting to the training data but then improved its performance on the validation set as it continued to train. This may have been due to a number of factors, such as changes in the model architecture, the use of regularization techniques, or the adjustment of hyper-parameters.

Overall, this pattern suggests that the model was able to improve its performance on the validation set over time, but it may still not be performing as desired.

5.2 Metalearner

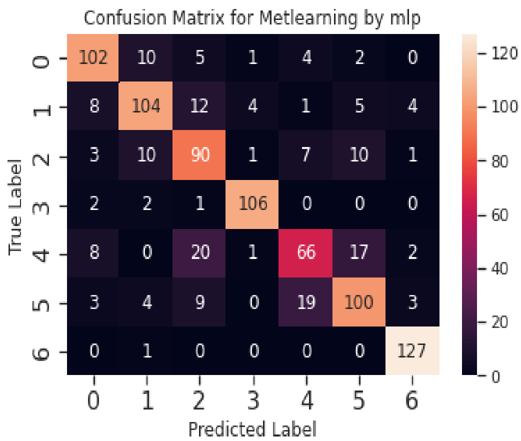


Figure 13: Confusion matrix of MLP

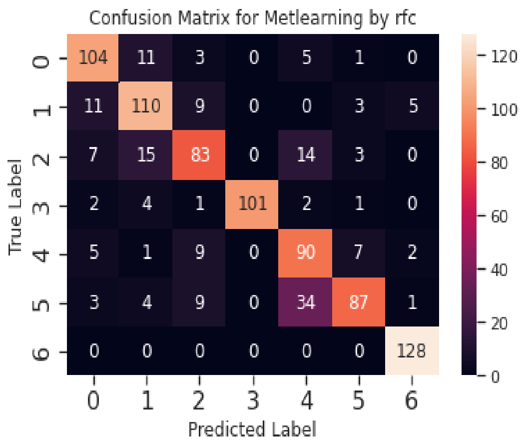


Figure 14: Confusion matrix of Random Forest

Here the model is developed by combining convolutional neural networks (CNN) and gray level co-occurrence matrix (GLCM) features to make predictions. The model first extracts features from images using a CNN, and then uses a GLCM to extract texture features from the images. These features are then combined and used as input to a meta learner, which is a machine learning model that uses the combined features to make predictions.

In this case, the meta learner is trained using three different algorithms: linear regression, multi-layer perceptron (MLP) and Random Forest. The linear regression model achieves an accuracy of 60%, while the MLP and Random Forest models both achieved an accuracy of 80%, Figure 13 and figure 14 show confusion matrix having large amount of True Positives and negligible number of type 1 and 2 errors which is more important for medical domain. This suggests that the MLP model is performing better than the linear regression model, likely due to its ability to learn complex non-linear relationships between the input features and the target variable.

Overall, this approach uses the strengths of both CNN and GLCM to extract relevant information from images and make accurate predictions. By combining the two methods and using a meta learner, the model is able to make more accurate predictions than being done alone.

5.3 Discussion

There are two main skin cancer classification models that have been developed. The first model uses a convolutional neural network (CNN) and has a loss of 1.1579, an accuracy of 56.73%, an F1 measure of 0.5401, a precision of 0.6557, and a recall of 0.4598. The accuracy graph for this model shows that the lines diverge until the 15th epoch and then overlap until the last epoch of 50.

The second model uses a combination of CNN features and gray-level co-occurrence matrix (GLCM) features that are fed into a meta-learner. The meta-learner uses 3 algorithms: logistic regression, Random Forest and multi-layer perceptron (MLP). The logistic regression algorithm gives an accuracy of 60%, while the other two algorithms gave an accuracy of 80%. , random forest gives precision of 81%, recall of 80.90% and f1 score of 0.8 which little over that of MLP algorithm. Table 5.1 illustrates evaluation all the suitable algorithhm used.

Table 5.1 Comparison of evaluation metrics

Metrics	CNN	MLP by Metalearner	Random Forest by Metalearner
Accuracy	56.49%	79.96%	80%
Precision	65.70%	77.70%	81.50%
Recall	46.43%	75.90%	80.90%
F1 Score	0.54	0.76	0.8

Overall, the second model having invloved MLP and Random forest appears to be more effective at skin cancer classification than the first model, as it achieves a higher accuracy and uses a meta-learning approach to improve performance.

6 Conclusion and Future Work

Author mainly developed ensembled model having included meta learner. There are two major steps. One being constructing CNN and another being development of metalearner. In the context of skin cancer detection, the performance metrics provided in the output would indicate that the CNN model is not very effective at identifying skin cancer. The low accuracy, F1 score, precision, and recall values suggest that the model is not making accurate predictions and may be missing a significant number of skin cancer cases. The use of a convolutional neural network along with meta-learning approach appears to be effective for skin cancer classification. The second model, which combines CNN and GLCM features and uses MLP algorithm, achieves a higher accuracy of 80%. Additionally, the use of image augmentation and Adam optimization improved the performance of the model.

Skin cancer is a serious and potentially life-threatening condition, so it is important for detection models to be as accurate as possible. A low-performing model could lead to missed diagnoses and delay in treatment, which could have serious consequences for patients.

In general, a good skin cancer detection model should have high accuracy, F1 score, precision, and recall values. This would indicate that the model is able to accurately identify a large number of skin cancer cases while also minimizing the number of false positives (incorrectly identifying healthy skin as cancerous). By achieving a high level of performance in these metrics, a skin cancer detection model can help to improve the speed and accuracy of diagnoses, ultimately leading to better outcomes for patients.

Future work involves in exploring the use of transfer learning or multi-task learning to improve the generalizability and adaptability of skin cancer classification models later on adding the layer of meta-learner could improve the performance exponentially. Current model only works on 2 dimensional data-set, Developing new skin cancer classification models that can handle a wider range of data types and image modalities, such as 3D images or multi-modal images that combine different types of imaging data. Investigating the potential of using unsupervised or semi-supervised learning methods to classify skin cancer, which could help to reduce the amount of labeled data required for training. Conducting further studies to evaluate the performance and clinical utility of skin cancer classification models in real-world settings.

Acknowledgement

I would like to express my deepest gratitude to my supervisor, Professor Noel Cosgrave, for his valuable guidance and support throughout my research. I am also grateful to the National College of Ireland for providing me with the necessary resources and support to conduct my research. I would also like to thank the participants who took part in my study for their valuable contributions. Finally, I would like to thank my family and friends for their constant encouragement and support.

References

Aguiar, G. J., Mantovani, R. G., Mastelini, S. M., de Carvalho, A. C., Campos, G. F. and Junior, S. B. (2019). A meta-learning approach for selecting image segmentation algorithm, *Pattern Recognition Letters* **128**: 480–487.
URL: *https://www.sciencedirect.com/science/article/pii/S0167865519302983*

Ajmal, A., Hollitt, C., Frean, M. and Al-Sahaf, H. (2018). A comparison of rgb and hsv colour spaces for visual attention models, *2018 International Conference on Image and Vision Computing New Zealand (IVCNZ)*, IEEE, pp. 1–6.

Ali, M. S., Miah, M. S., Haque, J., Rahman, M. M. and Islam, M. K. (2021). An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models, *Machine Learning with Applications* **5**: 100036.

Allugunti, V. R. (2022). A machine learning model for skin disease classification using convolution neural network, *International Journal of Computing, Programming and Database Management* **3**(1): 141–147.

Alquran, H., Qasmieh, I. A., Alqudah, A. M., Alhammouri, S., Alawneh, E., Abughazaleh, A. and Hasayen, F. (2017). The melanoma skin cancer detection and classification using support vector machine, *2017 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*, IEEE, pp. 1–5.

Barata, C. and Marques, J. S. (2019). Deep learning for skin cancer diagnosis with hierarchical architectures, *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, IEEE, pp. 841–845.

Dodge, S. and Karam, L. (2016). Understanding how image quality affects deep neural networks, *2016 eighth international conference on quality of multimedia experience (QoMEX)*, IEEE, pp. 1–6.

Fu, K., Zhang, T., Zhang, Y., Wang, Z. and Sun, X. (2021). Few-shot sar target classification via metalearning, *IEEE Transactions on Geoscience and Remote Sensing* **60**: 1–14.

Fu, K., Zhang, T., Zhang, Y., Yan, M., Chang, Z., Zhang, Z. and Sun, X. (2019). Meta-ssd: Towards fast adaptation for few-shot object detection with meta-learning, *IEEE Access* **7**: 77597–77606.

Guo, W., Xu, G., Liu, B. and Wang, Y. (2022). Hyperspectral image classification using cnn-enhanced multi-level haar wavelet features fusion network, *IEEE Geoscience and Remote Sensing Letters* **19**: 1–5.

Höhn, J., Krieghoff-Henning, E., Jutzi, T. B., von Kalle, C., Utikal, J. S., Meier, F., Gellrich, F. F., Hobelsberger, S., Hauschild, A., Schlager, J. G. et al. (2021). Combining cnn-based histologic whole slide image analysis and patient data to improve skin cancer classification, *European Journal of Cancer* **149**: 94–101.

Hoshyar, A. N., Al-Jumaily, A. and Sulaiman, R. (2011). Review on automatic early skin cancer detection, *2011 International Conference on Computer Science and Service System (CSSS)*, IEEE, pp. 4036–4039.

Huang, Z., Du, X., Chen, L., Li, Y., Liu, M., Chou, Y. and Jin, L. (2020). Convolutional neural network based on complex networks for brain tumor image classification with a modified activation function, *IEEE Access* **8**: 89281–89290.

Jain, S., Singhania, U., Tripathy, B., Nasr, E. A., Aboudaif, M. K. and Kamrani, A. K. (2021). Deep learning-based transfer learning for classification of skin cancer, *Sensors* **21**(23): 8142.

Lin, T.-C. and Lee, H.-C. (2020). Skin cancer dermoscopy images classification with meta data via deep learning ensemble, *2020 International Computer Symposium (ICS)*, pp. 237–241.

Mai, Z., Hu, G., Chen, D., Shen, F. and Shen, H. T. (2021). Metamixup: Learning adaptive interpolation policy of mixup with metalearning, *IEEE Transactions on Neural Networks and Learning Systems* .

Naeem, A., Farooq, M. S., Khelifi, A. and Abid, A. (2020). Malignant melanoma classification using deep learning: datasets, performance measurements, challenges and opportunities, *IEEE Access* **8**: 110575–110597.

Patel, V., Shah, S., Trivedi, H. and Naik, U. (2020). An analysis of lung tumor classification using svm and ann with glcm features, *Proceedings of First International Conference on Computing, Communications, and Cyber-Security (IC4S 2019)*, Springer, pp. 273–284.

Rezaoana, N., Hossain, M. S. and Andersson, K. (2020). Detection and classification of skin cancer by using a parallel cnn model, *2020 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE)*, IEEE, pp. 380–386.

Sedigh, P., Sadeghian, R. and Masouleh, M. T. (2019). Generating synthetic medical images by using gan to improve cnn performance in skin cancer classification, *2019 7th International Conference on Robotics and Mechatronics (ICRoM)*, IEEE, pp. 497–502.

Sharma, A. K., Tiwari, S., Aggarwal, G., Goenka, N., Kumar, A., Chakrabarti, P., Chakrabarti, T., Gono, R., Leonowicz, Z. and Jasiński, M. (2022). Dermatologist-level classification of skin cancer using cascaded ensembling of convolutional neural network and handcrafted features based deep neural network, *IEEE Access* **10**: 17920–17932.

Sultana, F., Sufian, A. and Dutta, P. (2018). Advancements in image classification using convolutional neural network, *2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)*, IEEE, pp. 122–129.

Xue, T. and Yu, H. (2020). Model-agnostic metalearning-based text-driven visual navigation model for unfamiliar tasks, *IEEE Access* **8**: 166742–166752.

Zhang, N., Cai, Y.-X., Wang, Y.-Y., Tian, Y.-T., Wang, X.-L. and Badami, B. (2020). Skin cancer diagnosis based on optimized convolutional neural network, *Artificial intelligence in medicine* **102**: 101756.