Drop-out Regularization: Enhancing Skin Cancer Classification with CNN

by

Nyi Nyein Aung

Sujung Kim

Final Report for IT 499

Payap University

Bachelor of Science in Information Technology

2023

I,....., hereby certify that this senior project report satisfies the final report requirements of IT 499.

_____ _____

"Signature" Date

Table of Contents

Table of Figures

Table of Tables

Abstract

The classification of skin cancers is a highly complicated process due to the variability and irregularities of skin conditions. Images taken by different devices often exhibit significant differences in dimensions, color saturation, and other factors, which aggravate the complexity of the classification process. A fast and accurate diagnosis is crucial in the field of dermatology to ensure patients receive timely treatment. Skin cancers can be automatically classified from images through the use of Convolutional Neural Network (CNN). This research aims to propose a Convolutional Neural Network Model with dropout regularization that can classify skin cancers through image classification. This approach enhances the adaptability and validity of the model, thus improving the accuracy of classifying dermatological diseases and mitigating the challenges posed by the variability of camera hardware. The experiment displays the effectiveness of the proposed method in achieving a good classification performance when tested on the skin cancer MINST HAM10000 dataset.

Acknowledgements

We would like to extend my heartfelt appreciation to those who have contributed significantly to this project. Our deepest thanks go to Dr. Supansa Chaising for her research idea that laid the groundwork for this project. We are immensely grateful to Dr. Wanus Srimaharaj for serving as my advisor and providing invaluable guidance that has been crucial in shaping this work. We are indebted to Grant Sanderson of the 3Blue1Brown YouTube channel for equipping us with amazing learning materials for image processing that greatly improved our understanding of this work. These individuals have been essential for the realization of this project and for their significant contributions, we are truly thankful.

1. Introduction

1.1 Background

The World Health Organization estimated that there were 325000 new cases of melanoma and 57,000 people died in 2020 due to melanoma of skin (IARC, 2023). The Skin cancer foundation predicts that every one out of five Americans will be diagnosed with skin cancer in their lifetimes (Skin Cancer Foundation, 2023). Skin cancer is a type of cancer that originates in the skin cells. It occurs when there is an uncontrollable growth of abnormal skin cells that can spread to other body parts and organs. There is a normal process for old skin cells to die off slowly and be replaced by new skin cells. However, after exposure to ultraviolet (UV) radiation from the sun or an artificial source like tanning beds, the rate at which the skin cells die fastens and starts to go out of control. The abnormal cell growth is considered 'benign' if it doesn't spread to other body parts but if it can spread, it is considered to be 'skin cancer'. This is the reason as to why ultraviolet radiation is one of the biggest factors to Skin Cancer. There are also other factors that contribute to skin cancer such as geographical location and skin type of the person.

There are three types of skin cancer: Basal cell carcinoma, squamous cell carcinoma and melanoma that are split into two categories: melanoma and non-melanoma. The Basel cell carcinoma and squamous cell carcinoma are classified as non-melanoma cancer while the melanoma cancer is classified as melanoma. Bascel Cell Carcinoma is the most common form of skin cancer and usually appears as a small shiny bump or nodule on the skin. The Squamous cell carcinoma often looks like a red, scaly patch that does not heal and sometimes spreads aggressively to other areas. Lastly, melanoma is the most dangerous form of skin cancer that develops from existing moles or appear as unusual growth on the skin with dark/brown color.

Skin cancer is the one of the most common forms of cancer globally as melanoma and non-melanoma skin cancers combined make up approximately one thirds of all cancers diagnosed globally according to the World Health Organization. It is crucial to have early detection for skin cancer, or any cancer in general, to achieve successful treatment and improve the chances of survival for the patient. Skin cancer is more likely to be treated effectively and suppressed when diagnosed at an early stage with much less aggressive interventions to the body. Consequently, decreasing the chances of the cancer spreading to other parts of the body. Skin cancer is often diagnosed through a combination of visual examination, dematoscopy and if deemed necessary by the treating physician, a skin biopsy. A skin biopsy is a process in which a small sample of abnormal skin tissue is taken from the patient to be tested under a microscope to determine if the cancerous cells are present. In recent times, the understanding of skin cancers' causes and treatments has significantly advanced due to scientific research and medical advances. One of the newer ways to diagnose skin cancer is using neural networks, specifically Convolutional Neural Network (CNN) to classify skin cancer automatically without the help of a physician.

Neural networks are a type of deep learning algorithm within the subset of machine learning that mimics the structure of the human brain. A neural network, also known as Artificial Neural Network (ANN), has three main layers: input layer, one or more hidden layer and output layer. Within those layers, there are nodes called 'neurons' that are connected to each other that have their own weights and biases. A neural network takes in a tensor of numbers into the input layers, passes that data through the hidden layer(s) and produces an output value at the output layer. It employs many mechanisms such as forward feedback and backpropagation during model training phase to adapt to the data and produce accurate predictions/classifications. There are many types of neural networks that are specialized to perform various tasks, I.e., Convolution neural

network (CNN) for image processing, large language model (LLM) for natural language processing and data generation.

Neural networks can be taught how to recognize diseases if there is enough training data. A patient's symptoms and other medical information are typically used by a physician to diagnose the patient's condition. The patterns from the decisions made by the physician as well as the characteristics of the patient can be recorded to be used for the development of an artificial neural network that can perform the task of the physician with an accuracy which is just as high. This will reduce the burden of the physician and healthcare staff whilst improving the experience of the patients and potentially saving more lives.

Convolutional neural networks are a type of deep learning neural network with a structure designed to tackle image processing and objective recognition jobs efficiently. It consists of the same layers as a typical neural network with the addition of the convolutional layer and a pooling layer. The convolutional layer takes the data of an image as the input and puts it through filters with variable number stride and padding. The filter kernel goes through the whole image extracting the feature information and outputting a feature map that will be passed to another convolutional layer or a pooling layer. The pooling layer reduces the amount of dimension of the data to reduce the parameters that need to be trained. Finally, the data will be sent to the ‘fully connected’ layer which has the same structure of the neural network described above that will perform the classification task and identify the object/thing from the input image.

CNNs can be implemented in medicine to detect various diseases including the type of skin cancer a patient has using images. CNN can take in the image of the patient, digest the features, and produce a diagnosis for the cancer type via image classification. The CNN model can be put inside a software that can be used by physicians and professionals to improve their diagnoses. It

can also be made into a program accessible to the public used to encourage early cancer detection. Although the accuracy might not be 100% and the diagnosis will not be medical grade, it can help improve people's awareness of cancers especially skin as the change is worsening and rates for skin cancers are rising(source).

This project aims to propose a CNN model with a drop out regularization technique. Drop-out Regularization is a computationally cheap technique to prevent a neural network from overlearning and becoming too complex by making neurons in the hidden layer inactive when a certain threshold is reached.

1.2 Objectives

1. Propose a new Convolutional Neural Network (CNN) model with a special Drop-out regularization technique
2. Revising and adjusting the CNN model architecture to apply it specifically to classifying skin cancer
3. Preprocessing the image data of the 10015 images of the Skin Cancer MNIST Ham10000 dataset
4. Achieve reasonable classification accuracy with validation accuracy equal or higher than 70%

1.3 Scope of Work

The project aims to build a Convolutional Neural Network (CNN) with a new Drop-out regularization technique that can classify skin cancer accurately by training the CNN model on the skin cancer MINST HAM10000 dataset that consists of 10015 images with 7 classes representing diagnostic cancer categories. The first part of the study will be focused on preparing the image dataset as training, validation, and testing data. The training, validation, and testing datasets will make up 64%, 16%, and 20% of the original dataset respectively. The proposed CNN model will

be trained with the training dataset while being validated by the validation dataset and the testing dataset will be used to evaluate the trained CNN model.

To split the images up into training and testing datasets, the HAM10000 metadata style sheet containing the class label of each image is used to match each image with a label before splitting the images into different folders. The task is done in a local environment on Jupyter Notebook using Python version 3.12. The Jupyter Notebook version 6.5.2 is hosted on a local server using the Anaconda Distribution. The current setup uses a CPU-only runtime for the execution of the task as splitting the images and moving them into specific folders do not require any complicated computation. Python libraries NumPy, Pandas, and Scikit-Learn are used to split the class labels into testing and training datasets randomly and fairly. Afterward, the OS and Shutil python libraries are used to create new directories and store the images from the dataset into specific directories to prepare the dataset to be used for the experiment.

The experiment section of the project is creating a CNN model and training it on the dataset to classify skin cancers. The first part of this section is to apply data augmentation techniques on the training dataset to perform normalization and to add randomness to the data which will improve the ability of the CNN model to learn the data. The proposed CNN model will be using dropout regularization as its main approach to overfitting. The model will be compiled to predict multiclass labels and fitted to the training dataset. After training the CNN model, its performance will be evaluated and validated using the testing dataset.

The experiment will be conducted offline on a local Python environment hosted via Anaconda Distribution and Visual Studio Code will be used as the main IDE for the experiment. The computer has the following specifications: 12th generation Intel Core i5-12500 H 2.50 GHz CPU, 16 GB of DDR5 RAM, and Nvidia RTX 3050 TI 4GB G6 GPU. TensorFlow version less

than 2.11 can be configured to directly use the GPU of the computer after installing CUDA on the local machine. This setup uses GPU runtime to optimize the training of the proposed CNN model. Python libraries such as TensorFlow, Scikit-Learn, and Kera's will be used in the experiment and TensorFlow's Sequential class will be used to design the CNN model architecture.

1.4 Plan and Schedule

Table 1. Schedule of the project

No	**Task**	**Schedule**
1	Planning	Week 1-2
2	Introduction	Week 3-4
2.1	Background on Skin Cancer	Week 3
2.2	Background on Convolutional Neural Network (CNN)	Week4
3	Literature Review	Week 5-6
3.1	Traditional Skin Disease diagnoses	Weel 5
3.2	Prior works on cancer classification using machine learning	Week 6
4	Methodology and experimenting	Week 7-12
4.1	Making a baseline CNN Model	Week 7
4.2	Experimenting to improve the Model	Week 8-11
4.3	Validate the final results	Week 12
5	Conclusion and finalizing	Week 13
6	Final Presentation	Week 14
7	Final Report	Week 15

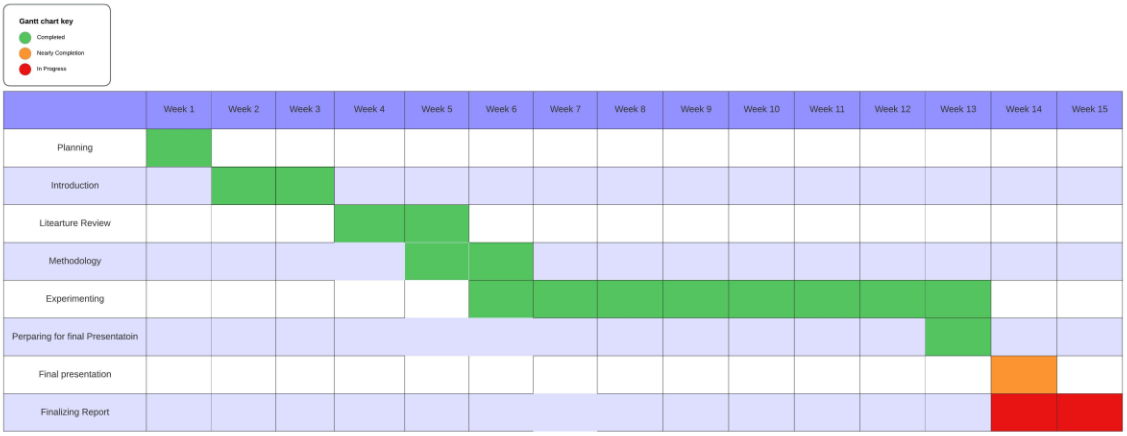

Figure 1. Gantt chart of the project

1.5 Conceptual Design

The proposed CNN model consists of 4 steps: Data Collection, Data Preprocessing, Model creation, and Model Testing shown in Figure 2.

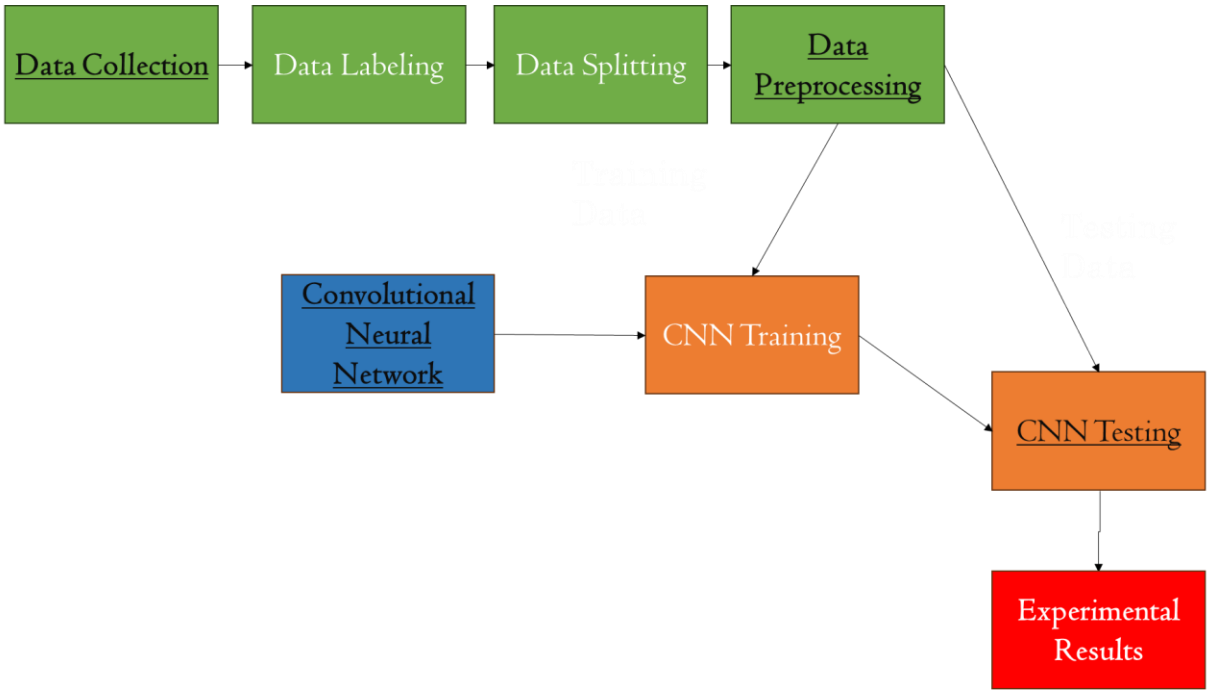

Figure 2. Conceptual Framework

2. Literature Review

2.1 Traditional Skin Disease and Skin Cancer Diagnoses

Skin diseases bear a significant amount of importance in the world of epidemiology as (Yakupu et al., 2019) found that 1.79% of the global disease burden was caused by skin diseases. The research also revealed that a total of 639,000 Disability-Adjusted Life Years (DALYs) and 2,013,000 Years of healthy life lost due to disability (YLDs) were caused by skin and subcutaneous diseases over the course of 41 years until 2013. This shows that skin diseases are a major cause of disability.

There are various skin diseases in the world ranging from dermatitis, psoriasis, impetigo, scabies, fungal skin diseases, viral skin diseases, malignant melanoma, ulcers, and squamous cell carcinoma. Globally, 34% of the identified cases of skin diseases were diagnosed as fungal skin diseases, and 23% were diagnosed as bacterial skin diseases (Yakupu et al., 2019). A total of 98.522 people died from skin and subcutaneous disease globally in 2019. Adding on to the huge variety of skin diseases, the location of the patient's homes and the weather conditions they face daily also have a huge factor for the type of skin disease that occurs. For example, most of the DALYs caused by skin and subcutaneous diseases were mainly condensed in China which accounted for 19% of the DALYs whereas melanoma was the cause reason for the largest disease burden in some regions such as Australia where white keratinocyte carcinoma was a major burden. There is also a noticeable difference between the sexes when diagnosed with skin and subcutaneous diseases. Although men and women had fungal skin diseases as the most common skin diseases men's age-standardized incident rate (ASIR) for fungal diseases was higher than women's at 22% (Yakupu et al., 2019).

According to the World Cancer Research Fund International (WCRF, 2020), as of 2020, non-melanoma skin cancer rates around the world are over three times the rate of melanoma skin cancer with approximately 1.2 million cases compared to around 325 thousand cases. Australia had the highest overall non-melanoma cancer rate around the world with 91,000 cases and the highest melanoma cancer rate with 25,000 cases in 2020. A variety of factors are associated with skin cancer but sunlight and specifically long exposure to UV light makes humans more susceptible to both melanoma and non-melanoma skin cancer like the case with Australia. The general dermatologic diagnosis typically involves two steps which are 1) observation through physical examination, clinician investigations, and taking the patient's history and 2) the interpretation of gathered information in terms of functionality, structural, and pathological disorders (Mahajan V.K. & Handa S., 2020). Coincidently, a huge number of dermatologic cases are instantly diagnosable with a simple look through direct inspection. Several factors also play a role in the diagnosis such as racial, geographical, and environmental differences. Therefore, a detailed medical history with good investigative work is very important for arriving at an accurate diagnosis. (Coulson et al. 2016; Garg et al. 2008; Pasricha and Khaitan 2005) produced a general outline for history taking and examination in dermatology shown in Table 2.

Table 2. General outline for history taking and examination in dermatology (Coulson et al. 2016; Garg et al. 2008; Pasricha and Khaitan 2005)

Chief or present complaint and treatment(s)	• Duration, continuity, or intermittency • Evolution (improvement, worsening, or appearance of new lesions) • Symptoms such as itching, pain, burning, sensory-motor loss, etc.

	• Diurnal variation of symptoms • Past episode (if positive history for past episodes, or similar presentation) • Investigations, treatments, and their effects done for the present condition
Past medical history	• Past or concurrent illnesses such as diabetes, hypertension, etc. • Other skin diseases • Therapies
Family history	• Similar diseases • Other diseases such as asthma, melanoma, genetic disorders, etc. • Consanguinity of parents
Sexual practices (optional)	• In patients with sexually transmitted diseases (STDs), it is needed
Menstrual/obstetric history in women	• Menstrual irregularities • Bad obstetric history
Personal and social history, occupation, hobbies and spare time activities	• Habits such as sleep, diet, alcohol consumption, smoking, recreational drugs • Intake of vitamin and other supplements, pain killers, laxatives, and other treatments • Exposure to chemicals, cement, and weeds • Hobbies • Travels
Physical examination	• General appearance including obesity, height, mental and physical orientation, etc.

	• Vitals, temperature, blood pressure, paleness, cyanosis, etc. • General mucocutaneous examination like skin, hair, nails, etc. • Specific mucocutaneous examination related to the dermatologic complaint

Table 1 shows the outline of the details that a practicing physician and the nurse have to gather from a patient for every single diagnosis. They also have to employ a variety of skin examination tools to closely analyze the skin lesions or any problem the patient has concerns about. A few of the tools include 1) a diascope for diascopy to highlight the structure of a lesion with a vascular component, 2) a wood lamp that emits UV light to make certain types of cells turn colors, 3) a dermatoscope to examine the skin using a high-magnification microscope and 4) Trischscope for close visualization of hair and scalp with 20x to 120x magnifications.

The dermatological (skin and subcutaneous) diseases also have to be evaluated to arrive at a diagnosis like most diagnoses. However, the methods employed tend to be subjective and might lead to discrepancy in results which leads to some causes that cannot be reproduced due to interobserver variation (Mahajan V.K. & Handa S., 2020). Table 3 shows uniformly accepted and commonly used scoring systems in dermatology used for diagnostic purposes.

Table 3. Commonly used scoring systems in dermatology (Mahajan V.K. & Handa S., 2020)

No	Name	Description
1	Wallace Rule of Nine	The skin area of the palm of one hand is equivalent to 1% of the total body surface area (BSA). The Wallace rule of nine is used to measure the involved BSA to estimate the severity of widespread skin diseases

2	Psoriasis Area Severity Index Score (PASI Score)	Used for assessment of extent and severity of psoriasis despite limitation of interobserver variability by combining assessment of psoriasis-induced erythema, scaling, and skin thickness by weighting them according to the size of the affected area (Fredriksson and Pettersson, 1978)
3	Melasma Area Severity Index (MASI)	Used for measuring severity of melasma before and after treatment t (Kimbrough-Green et al. 1994)
4	Scoring for Vitiligo	Vitiligo area severity index (VASI) and vitiligo disease activity (VIDA) score are two systems used to access disease activity and severity in Vitiligo (Hamzavi et al., 2004)
5	Scoring Atopic dermatitis (SCORAD)	Used to standardize the assessment of severity and interpret therapeutic outcome in atopic dermatitis despite interobserver variability (Stalder et al., 1993)
6	Urticaria Activity Scores (UAS)	A patient-reported measure designed fo chronic spontaneous urticaria that comprises of the wheal number score and the itch severity score. The USA 7 is the sum of the daily average UAS over 7 days (Zuberbier et al., 2014)
7	Score for Toxic Epidermal Necrolysis (SCORTEN)	Used to access prognosis and severity in patients with toxic epidermal necrolysis (TEN)

A common theme found in the evaluation methods for typical skin diseases including skin cancers shown in Table 3 is the problem with interobserver variability. Interobserver variability is hard to control especially for skin and subcutaneous diseases because despite having similar

evidence, the eyes of the observers might not necessarily be seeing the same thing as one another. This will lead to problems in producing a conclusive diagnosis that covers all the ground easily. Fast and accurate diagnoses are very important for patients as skin cancer needs to be treated as fast as possible to prevent uncontrollable destructive consequences. Therefore, in recent years, medical professionals have started employing artificial intelligence through the use of neural networks to perform simple classifications of diseases to tackle the problem of interobserver variability.

2.2 Prior works on Skin classification using ML, NNs and CNNs

There have been various studies in the past decade that have been attempting to automate the diagnostic process of skin skin cancers. Alam et al. (2016) used support vector machine on an image with feature selection using texture-based information to automate the evaluation of the eczema process. Typical machine learning approaches such as Bayesian classification, Decision tree classification, and K-Nearest Neighbors are not appropriate for skin cancer classification as they are not suitable to handle such large amounts of data and the zero-probability problem plays a huge factor in their ability to be implemented properly. Verma et al. (2019) found that the ensemble models produce a higher accuracy for skin disease classification compared to traditional classification models. Lee et al. (2018) has achieved practical validation accuracy of 89.90% with their Fine-Tuned Neural Network skin disease classification model but stated that the network can be calibrated to higher accuracy with significant effort.

The Convolutional Neural Network (CNN) is one of the most prominent methods for skin cancer classification as there are studies like (Harangi, B., 2018) that show promising results after employing popular CNN approaches like AlexNet and VGGNet architectures. Despite the promising results, employing the CNN approach in the real-world setting produces quite a difficult

problem as it is a challenging task to get the model working well with data from different mobile phone cameras and digital cameras.

A few studies like (Yao et al., 2022) applied regularization techniques like the Dropout technique and image augmentation in conjunction with their novel loss function to produce a model with high skin lesion classification while others like (Padmavathi et al., 2022) applied transfer learning methods that involved using a pre-trained deep learning neural network and fine-tuned the network's components to produce a high-performance model for classification, ultimately to tackle the application problem mentioned before.

The model proposed by this study will also combine the approaches used in (Yao et al., 2022 & Padmavathi et al., 2022) by employing various pre-trained models using transfer learning and adding the Dropout regularization technique along with image augmentation. The proposed method will be designed to tackle the problem of camera lens variability and specifications which is essentially the same problem as the traditional methods of skin disease classification faced previously with interobserver variability. Applying transfer learning is another benefit to the implementation of the CNN as the computational cost of training the model on very large sets of data is no longer the burden of the experiment which helps save costs and improves the chances of reproducibility.

3. Methodology

3.1 Theorical Framework

3.1.1 Convolutional neural networks (CNNs)

Convolutional neural networks are typically used for image classification tasks due to their superior performance with image, speech, or audio signal inputs. A CNN is made up of special layers called the convolution layer followed by the pooling layers.

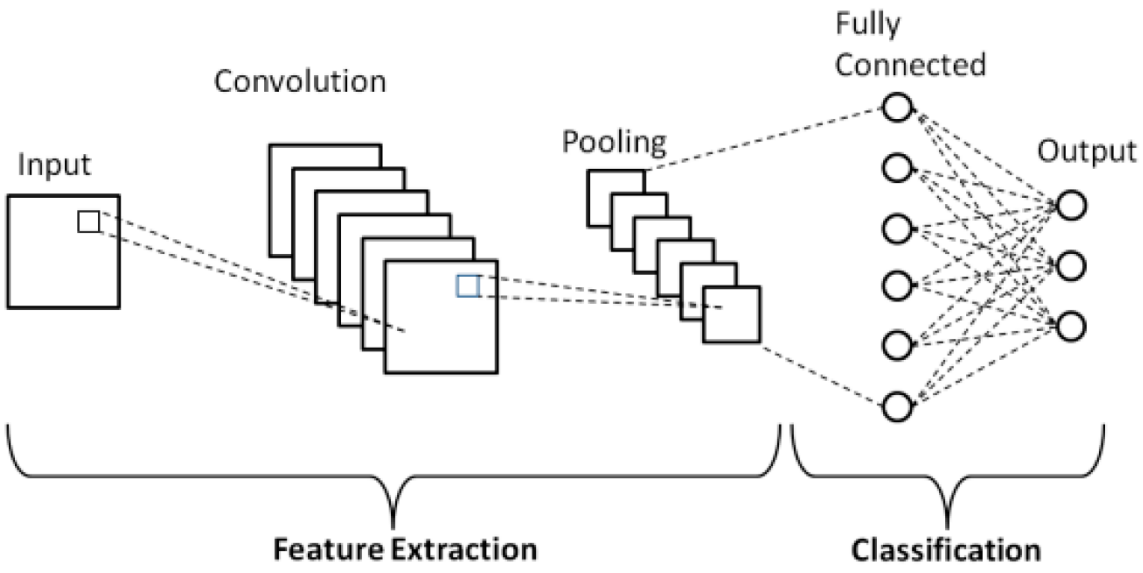

Figure 3. Structure of a convolutional neural network (Phung VH & Rhee EJ, 2019)

3.1.2 Convolutional layer

The convolutional layer is a fundamental part of a CNN as it captures the features and patterns in the input data of the images. The layer has learnable filters or kernels which are a small grid of weights, and they slide over the input data to perform the convolution operation by calculating the element-wise multiplication at each position and summing the results. This filtering process allows the neural network to recognize specific patterns such as textures and edges and represent them in the output known as the feature maps, making it crucial for processes like image recognition. The filters There can be multiple filters stacked on the input data to detect several different features.

Padding and Strides are two important hyperparameters of the convolutional layer that influence the convolution process. Padding is a process that involves adding additional rows and columns of zeros to the input data before the convolution to preserve the spatial dimensions of the data by making sure the features on the edges are not lost.

Strides, on the other hand, determine the size of the step that the kernels slide over the input data. A larger stride reduces the spatial dimension of the output feature map whereas a smaller stride preserves more spatial dimensions. These two processes are important for controlling the size of the feature maps in the convolutional layer and subsequently the output. Figure 3 shows an example way a stride can be visualized, and Figure shows how padding is applied during the filtering process.

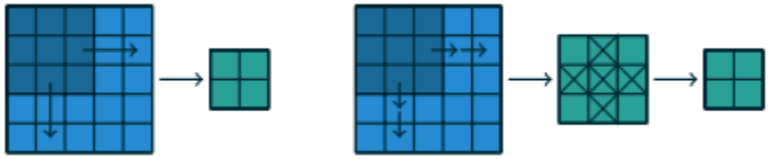

Figure 4. An example way to view strides (Dumoulin, V., & Visin, F., 2016)

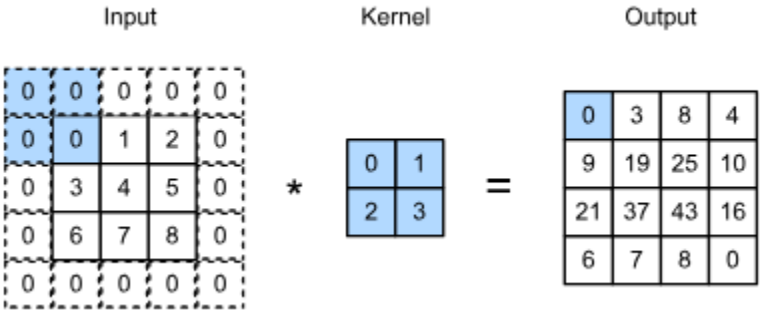

Figure 5. Example view of padding during the filtering process (D2l.ai., n.d.)

3.1.3 Pooling layer

The pooling layer is a component of the Convolutional Neural Networks (CNN) that is responsible for reducing the spatial dimensions of the feature maps while retaining important information. This layer is crucial for reducing the computational costs of the CNNs and improving the networks' efficiency during the training process. The pooling layer applies a pooling operation to a region of the feature map to perform its function.

Max pooling is the most common pooling operation that works by sliding a rectangle window typically sized 2x2, 3x3, 5x5, and 7x7 across the feature map. At each position, the maximum value of the window is selected and retained in a down-sampled feature map. Max pooling works by preserving the most influential feature within each region during the down-sampling process.

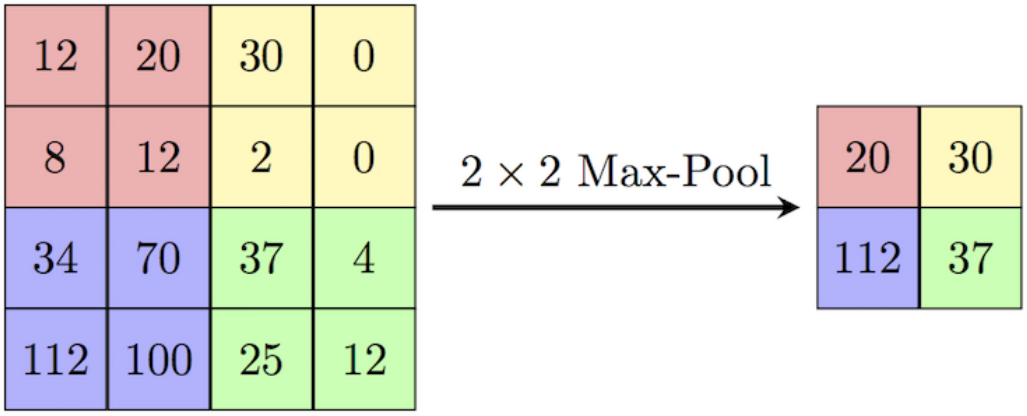

Figure 6. Max pooling operation example

In Figure 6, each shade of color represents the area of effect of each pooling operation done using max pooling. For instance, a 2 by 2 square covers over 4 pixels of data: 12,8,20, 12 and selects the highest value of 20 to represent this area in the new feature map shown on the right.

3.1.4 Fully Connected Layer

The fully connected layer, also known as the dense layer, is a component of the Convolutional Neural Network responsible for the classification task. It is located at the end of a CNN after the convolutional and pooling layers to interpret the highly detailed feature maps and predict the class of the output. The layer takes in the weighted sum of

inputs in the previous layer and passes it through an activation function that typically follows non-linearity, for example, the rectified linear unit (ReLU) activation function. The final layer of the fully connected layer works as the output layer and uses the SoftMax function to predict the probability scores of outputs in the classification problems. In a sense, the fully connected layer is like a traditional neural network that uses feature maps as input data to perform predictions. The type and amount of layers it can contain can be altered to obtain the best results whilst preventing overfitting, i.e., drop-out layer.

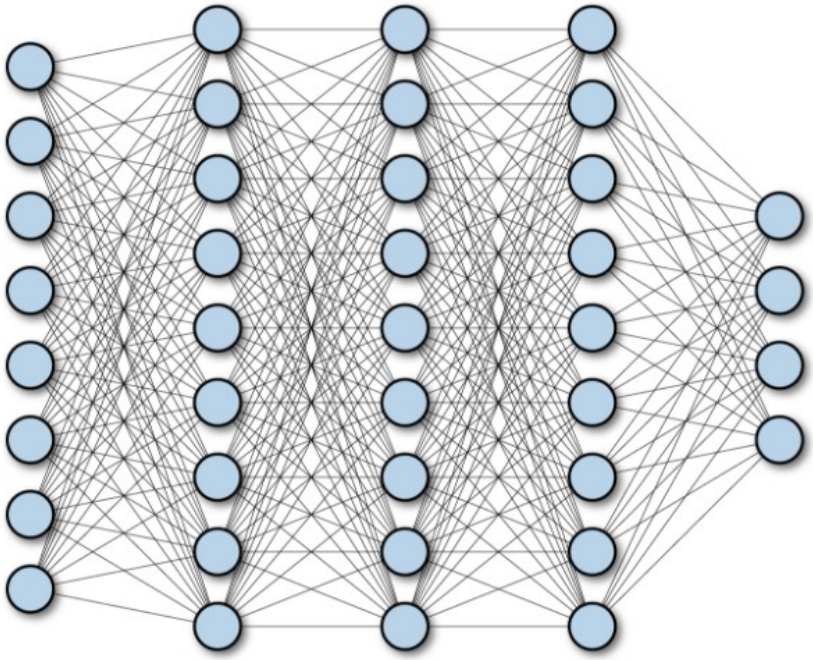

Figure 7. Example Fully Connected Layer

3.2 Data and Method

3.2.1 Data collection

This study is using the skin cancer MNIST Ham10000 dataset from Kaggle (Tschandl et al, 2018). This dataset contains 10015 images that comprises of 7 cancer diagnostic categories:

1. Actinic keratoses and intraepithelial carcinoma / Bowen's disease (akiec)
2. Basal cell carcinoma (bcc),
3. Benign keratosis-like lesions (solar lentigines / seborrheic keratoses and lichen-planus like keratoses, bkl)
4. Dermatofibroma (df)
5. Melanoma (mel)
6. Melanocytic nevi (nv)
7. Vascular lesions (angiomas, angiokeratomas, pyogenic granulomas and hemorrhage, vasc).

These seven classes cover about 95% of all pigmented lesions examined in the clinical practice of the two study sites. The data was collected over 9 years from 2008 until 2018 and 50% or more of the lesions included in the dataset were verified by pathology and the ground truth for the remainder of the cases was either follow-up, expert consensus, or confirmation by in-vivo confocal microscopy. The ethical review committee of the Medical University of Vienna and the University of Queensland approved the data collection, and the data of the patients were anonymized to the best of the researcher's knowledge. 64% of the data will be used as a training dataset, 16% for the validation dataset, and 20% testing dataset. The original class distribution of the skin cancer MINST HAM10000 dataset is shown in Figure 8.

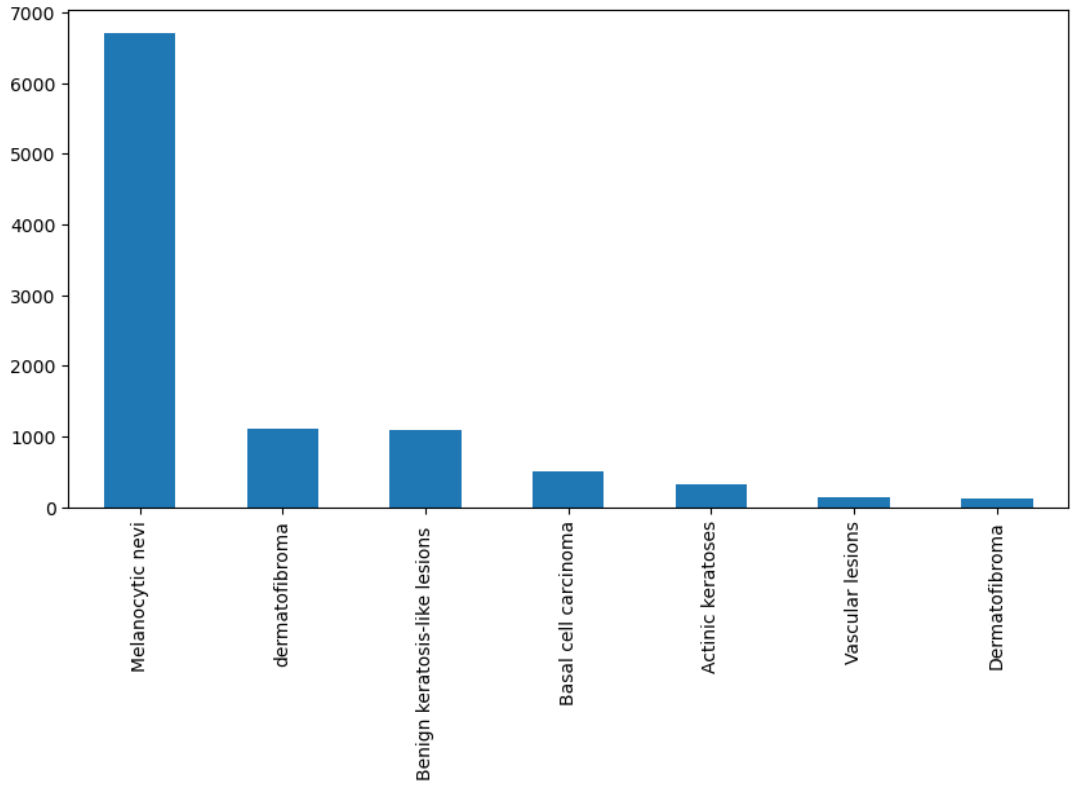

Figure 8. Class distribution of Cancer categories

3.2.2 Image Augmentation

This study employs 5 preprocessing steps on the image data of which, 4 are data augmentation techniques. The first step is the rescaling technique where the pixel data of each image is scaled to a range of 0 to -1 by dividing the color value of each color channel of each pixel as shown in Equation 1.

$$Y = \frac{X}{255} \tag{1}$$

Where Y is the rescaled pixel and X is the original X value.

Rescaling is applied to the training data as well as the validation and the testing data to normalize the standard of the input data. The other 4 data augmentation techniques are

random rotation, random horizontal flip, random shear, and random zoom. Random rotation rotates the input images -10 to 10 degrees randomly. Random horizontal flip, as the name suggests, flips the image horizontally with a probability of 10%. Random shear slants the image randomly by +/- 10 degrees randomly. Lastly, random zoom enlarges the image by up to 10%. These techniques are applied to the training dataset during the training phase to introduce randomness to the data which will help prevent the model from overfitting.

3.2.3 Proposed CNN Model Architecture

As shown in table 4, the proposed CNN model architecture contains 11 layers, including an input layer, the convolutional section that consists of three pairs of convolutional layers and a max pooling layer followed by a drop-out layer, the fully connected layer that contains the flatten layer, a Dense layer and a dropout layer, and the output layer.

Table 4. Proposed CNN Model Architecture Summary

Layer Type	**Output Shape**	**Params**
Conv2D	(None, 30, 30, 256)	7168
MaxPooling2D	(None, 15, 15, 256)	0
Conv2D	(None, 13, 13, 128)	295040
MaxPooling2D	(None, 6, 6, 128)	0
Conv2D	(None, 4, 4, 64)	73792
MaxPooling2D	(None, 2, 2, 64)	0
Dropout	(None, 2, 2, 64)	0
Flatten	(None, 256)	0
Dense	(None, 32)	8224
Dropout	(None, 32)	0

Output Layer	(None, 7)	321
Total params: 384,455 Trainable params: 384,455 Non-trainable params: 0		

This CNN model took in image data with the input shape of (32,32,3) which represents the image width, image height and number of color channels.

In the Convolutional section, the first convolutional layers had a kernel size of 3 x 3, used no padding and employed the ReLU activation function. The number of filters is 256, 128 and 64 in their respective order. The max pooling layer had a pooling size of 2 by 2 pixels.

In the fully connected layer section, the Dropout layers had an occurrence chance of 30% during training. The Dense layer in the fully connected layer has 32 neurons and uses the ReLU activation function. Lastly, the final output layer produces the likelihood of the class diagnosis of the image data through the SoftMax activation function.

3.2.4 Model Training and Testing

The proposed CNN model is fitted with the Adam optimizer with a learning rate of 0.001 and trained using the Sparse Categorical Cross Entropy loss function. The model is trained for 36 epochs as an early stopper monitoring the difference between the validation loss and the training loss stopped after the model stopped showing progress. The model with the best weights will then be saved and evaluated using accuracy, precision, f1 score, and recall metrics using the testing dataset. A confusion matrix of the proposed CNN model's predictions will also be used to analyze the performance of the model on the classes. A ROC-curve graph will also be employed to analyze the classification strength of the model on different cancer categories.

4. Experimental Results

The experiments produced promising results that demonstrate the effectiveness of the proposed CNN data in classification of cancer categories using image classification. Table 5 shows the classification performance of the proposed CNN model on the testing data. The models show good performance by obtaining the accuracy of 75%. This is further supported by the average weighted values of precision, recall and f1-core which are 0.72, 0.75 and 0.73 respectively. The means that the proposed CNN model is able to successfully classify the cancer category of an image 75% of the time with a precision of 72% which is a slightly better performance than the study objective 4. The unique thing to note is that the precision, recall and f1-score of the Bascel Cell Carcinoma class were 0.00 which shows that the model didn't predict any of this category correctly.

Table 5. The classification performance of the proposed CNN model

Class	Precision	Recall	F1-score	Accuracy
Melanocytic nevi	0.41	0.13	0.2	-
Dermatofibroma	0.55	0.43	0.48	-
Benign Keratosis-like lesions	0.47	0.41	0.44	-
Basal Cell Carcinoma	0.00	0.00	0.00	-
Actinic keratoses	0.57	0.38	0.45	-
Vascular lesions	0.82	0.94	0.88	-
Dermatofibroma	0.58	0.72	0.64	-
Weighted Average	0.72	0.75	0.73	75%

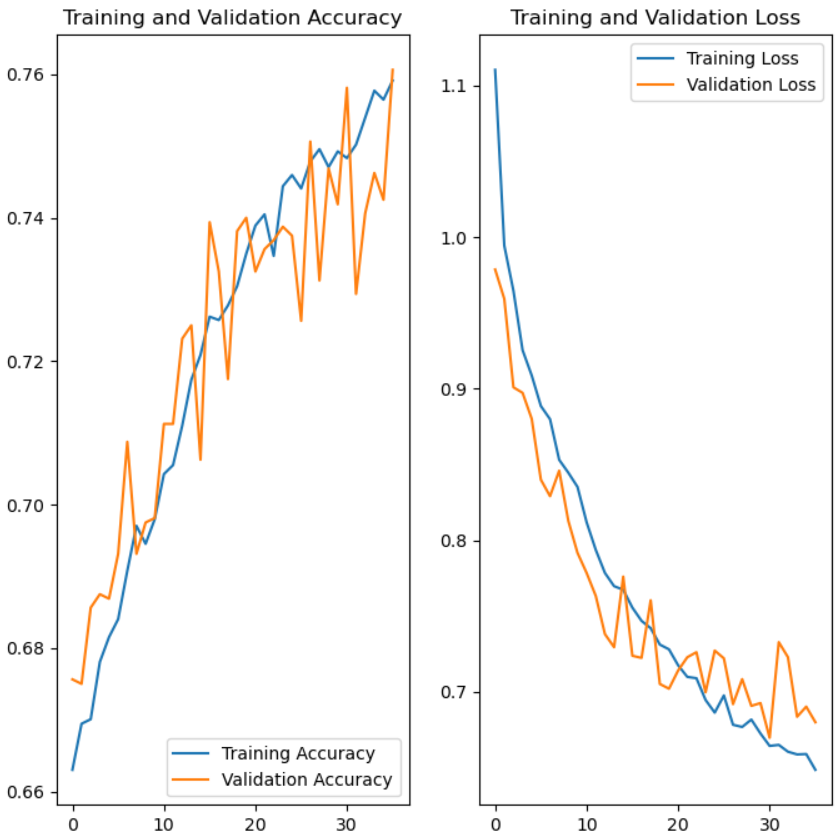

Figure 9. The training and validation history graphs of the proposed CNN model

Figure 9 displays the training history of the proposed CNN model which shows that the validation accuracy and the training accuracy didn't have any significant differences throughout the 36 epochs of training. The training and validation accuracy seems to be at a similar percentage during the training period, but the training loss starts to be significantly lower than the validation accuracy after the 31 epochs which is when the early stopper callback ended the model training to prevent overfitting. Overall, from looking at the training and validation histories, the proposed CNN doesn't show any signs of overfitting to the data and shows good robustness and generalization.

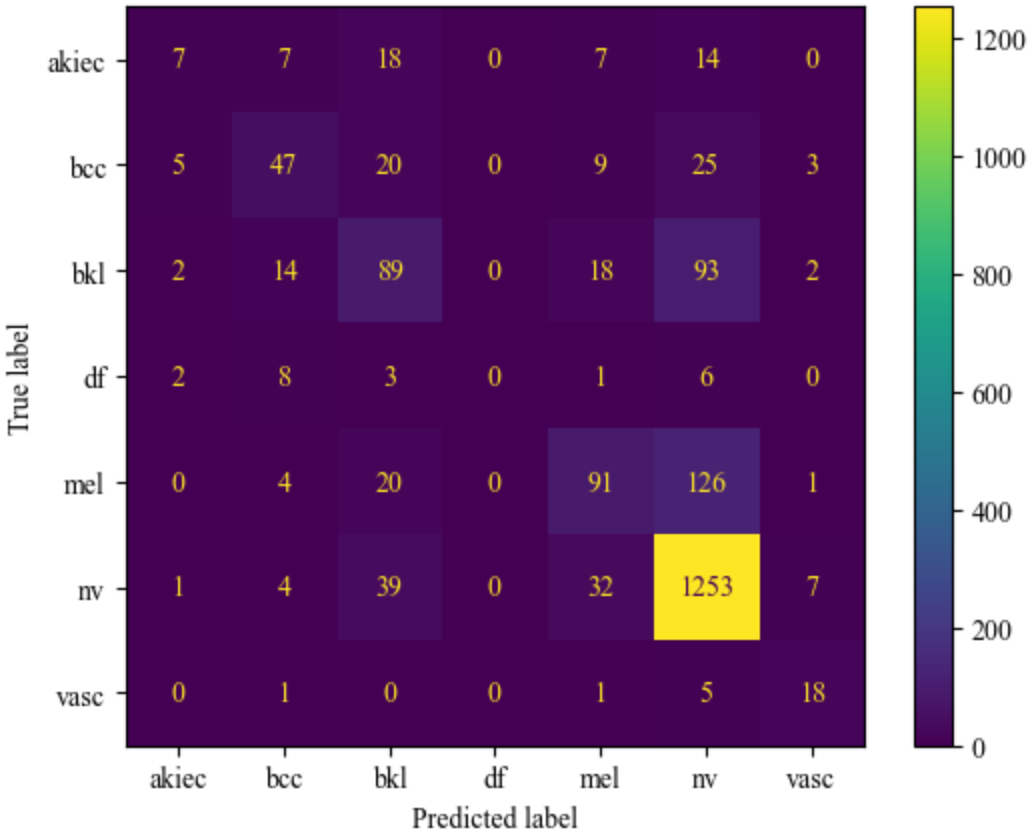

Figure 10. Confusion Matrix of the proposed CNN model

Figure 10 shows the confusion of the proposed CNN model. It shows that the model can predict most of the Melanocytic Nevi class correctly, but it also incorrectly predicted a few of the true instances of the Melanocytic Nevi class as Melanoma and Benign keratosis-like lesions classes. Overall, the confusion matrix exhibits a consistent pattern of correct performances with sight deviations which indicates a good performance.

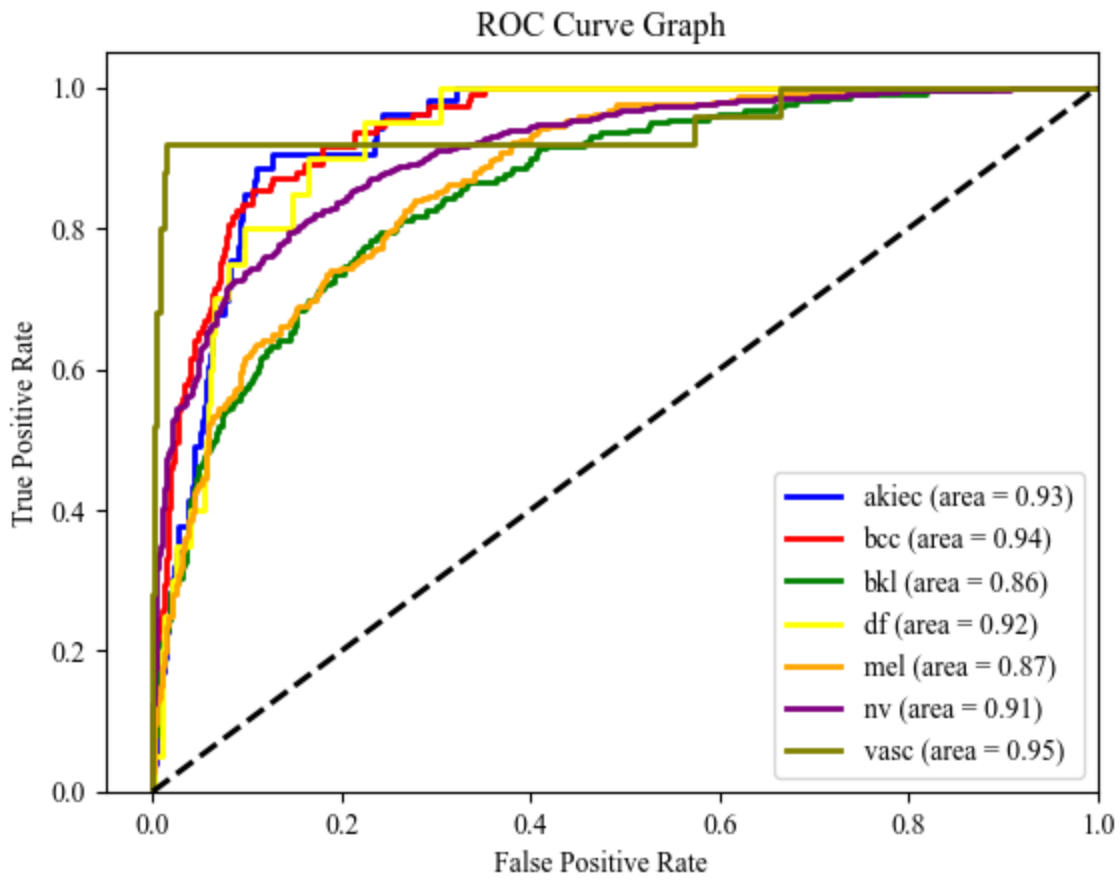

Figure 11. Multiclass ROC Curve Graph of the Proposed CNN Model

Figure 11 displays the multiclass ROC curve graph of the proposed model and the ROC curves of each class. All of the classes achieved an AUC value equal to or higher than 0.86 and as high as 0.95. This shows that the model can distinguish between the 7 classes properly with good performance. The vascular lesions class had the highest AUC score of 0.95 which indicates that the model can classify this class 95% of the time correctly. The class with the lowest AUC score was the benign keratosis-like lesions with a score of 0.86 which still indicates a strong performance. The ROC curve graph shows no biases towards the majority class which shows the robustness of the proposed CNN model.

5. Conclusion

Skin diseases represent a major medical illness that affects the skin's appearance, texture, color, and overall health. There are millions of identified cases of skin diseases annually causing substantial disability-adjusted life years (DALYs) and years lived with disability (YLDS to the population. There is a huge variety of skin disease types that have different prevalence in different regions due to various factors that cause skin diseases like genetics, environment, and gender. Skin cancer is one of the most prevalent and deadly types of skin disease. The overexposure to sunlight and UV radiation makes humans more susceptible to skin cancers. Melanoma and non-melanoma skin cancers are two main forms of skin cancer, but the latter isn't well documented due to the majority of cases being treated successfully. This means that the instances of skin cancer might even be underreported globally.

Skin diseases have traditionally been diagnosed by medical doctors based on standard scoring systems after examining the visiting patient's health condition and analyzing their medical history. This process suffers from interobserver variability as the diagnosis is highly subjective to the doctor's judgment at the time of the consultation. In recent years, efforts to automate the diagnostic processes in the medical field using machine learning have become increasingly popular and in the case of skin diseases, image classification is leading the way.

Most notably, the Convolutional Neural Network (CNN) has shown the most promise in diagnosing skin diseases automatically using images thanks to the convolutional layers giving the CNN strong capability for image classification. CNN can improve the supply-demand gap of dermatologists in many parts of the world and improve the accessibility of skin disease diagnoses. As melanoma is one of the main types of skin diseases, early detection is essential for any patient,

the newer method can enhance the diagnostic speeds drastically given that they can produce accurate diagnoses.

Therefore, this study aimed to propose a CNN model with dropout regularization to classify skin cancer categories using image classification. The study was able to create a CNN model with an average testing and validation accuracy of 75% and an average precision of 75%. The proposed CNN model didn't display any biases towards the majority class in the testing dataset which shows strong robustness and generalization of the model. The results indicate that the proposed CNN is a useful method for classifying different categories of skin cancer using image classification.

6. Future Studies

In the future, more studies with larger and more balanced datasets could improve the classification performance of the proposed CNN model but the class distributions of the dataset should be somewhat representative of the real-world clinical situations. Data augmentation techniques can also be explored further during the preprocessing phase as an additional regularization technique to the dropout technique to allow for the creation of deeper models with longer epochs of training.

References

1 Alam, M. N., Munia, T. T. K., Tavakolian, K., Vasefi, F., MacKinnon, N., & Fazel-Rezai, R. (2016, August). Automatic detection and severity measurement of eczema using image processing. In 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 1365-1368). IEEE.

2 Coulson, I. H., Benton, E. C., & Ogden, S. (2016). Diagnosis of skin disease. Rook's Textbook of Dermatology, Ninth Edition, 1-28. pp 4.19–4.22

3 D2l.ai. (n.d.). Padding and Strides. Dive into Deep Learning. https://d2l.ai/chapter_convolutional-neural-networks/padding-and-strides.html

4 Dumoulin, V., & Visin, F. (2016). A guide to convolution arithmetic for deep learning. arXiv preprint arXiv:1603.07285.

5 Fredriksson T, Pettersson U (1978) Severe psoriasis-oral therapy with a new retinoid. Dermatologica 157(4):238–244

6 Garg A, Levin NA, Bernhard JD (2008) Structure of skin lesions and fundamentals of clinical diagnosis. In: Wolff K, Goldsmith LA, KatzSI, Gilchrest BA, Patter AS, Leffell DJ (eds) Fitzpatrick's textbook of dermatology in general medicine, 7th edn. Mc-Graw-Hill, New York, pp 26–40

7 Groh, M., Harris, C., Daneshjou, R., Badri, O., & Koochek, A. (2022). Towards transparency in dermatology image datasets with skin tone annotations by experts, crowds, and an algorithm. Proceedings of the ACM on Human-Computer Interaction, 6(CSCW2), 1-26.

8 Hamzavi, I., Jain, H., McLean, D., Shapiro, J., Zeng, H., & Lui, H. (2004). Parametric modeling of narrowband UV-B phototherapy for vitiligo using a novel quantitative tool: the Vitiligo Area Scoring Index. Archives of Dermatology, 140(6), 677-683.

9 Harangi, B. (2018). Skin lesion classification with ensembles of deep convolutional neural networks. Journal of biomedical informatics, 86, 25-32.

10 IARC. (2023). Skin Cancer. World Health Organization. https://www.iarc.who.int/cancer-type/skin-cancer/

11 Kimbrough-Green CK, Griffiths CE, Finkel LJ, Hamilton TA, Bulengo-Ransby SM, Ellis CN, Voorhees JJ (1994) Topical retinoic acid(tretinoin) for melasma in black patients. Arch Dermatol 130(6):727–733

12 Lee, Y. C., Jung, S. H., & Won, H. H. (2018). WonDerM: Skin lesion classification with fine-tuned neural networks. arXiv preprint arXiv:1808.03426.

13 Mahajan V.K., Handa S. (2020) Diagnosis of Skin Diseases. In: Smoller B., Bagherani N. (eds) Atlas of Dermatology, Dermatopathology and Venereology, 1st edn. Springer, Cham. DOI: https://doi.org/10.1007/978-3-319-45134-3_9-1

14 Padmavathi, K., Neelam, H., Reddy, M. P. K., Yadlapalli, P., Veerella, K. S., & Pampari, K. (2022, January). Melanoma Detection Using Deep Learning. In 2022 International Conference on Computer Communication and Informatics (ICCCI) (pp. 1-4). IEEE.

15 Pasricha JS, Khaitan BK (2005) Clinical methods in dermatology and venereology, 5th edn. Mehta Offset Pvt Ltd, New Delhi

16 Phung VH, Rhee EJ. A High-Accuracy Model Average Ensemble of Convolutional Neural Networks for Classification of Cloud Image Patches on Small Datasets. Applied Sciences. 2019; 9(21):4500. https://doi.org/10.3390/app9214500

17 Skin Cancer Foundation. (2023a, March 6). Skin cancer facts & statistics. The Skin Cancer Foundation. https://www.skincancer.org/skin-cancer-information/skin-cancer-facts

18 Stalder JF, Taïeb A, Atherton DJ, Bieber P, Bonifazi E, Broberg A, Calza A, Coleman A, De Prost Y, Stalder JF, Gelmetti C, Cuannetti A, Harper J, Künz B, Lachapelle JM, Langeland T, Lever R, Oranje AP, Oueille-Roussel C, Revuz J, Ring J, Roujeau JC, Saurat JH, Song M, Tennstedt D, Van Neste D, Vieluf D, Poncet M (1993) Severity scoring of atopic dermatitis: the SCORAD index: consensus report of the European task force on atopic dermatitis. Dermatology 186(1):23–31

19 Tschandl, P., Rosendahl, C., & Kittler, H. (2018). The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific data, 5(1), 1-9.

20 Verma, A.K.; Pal, S.; Kumar, S. Classification of Skin Disease using Ensemble Data Mining Techniques. Asian Pac. J. Cancer Prev. 2019, 20, 1887–1894.

21 WCRF. (2022). Skin cancer statistics: World cancer research fund international. Retrieved from https://www.wcrf.org/cancer-trends/skin-cancer-statistics/

22 Yao, P., Shen, S., Xu, M., Liu, P., Zhang, F., Xing, J., ... & Xu, R. X. (2021). Single model deep learning on imbalanced small datasets for skin lesion classification. IEEE transactions on medical imaging, 41(5), 1242-1254.

23 Yakupu, A., Aimaier, R., Yuan, B., Chen, B., Cheng, J., Zhao, Y., Peng, Y., Dong, J., & Lu, S. (2023). The burden of skin and subcutaneous diseases: findings from the global burden of disease study 2019. Frontiers in public health, 11, 1145513. https://doi.org/10.3389/fpubh.2023.1145513

24 Zuberbier T, Aberer W, Asero R, Bindslev-Jensen C, Brzoza Z,Canonica GW, Church MK, Ensina LF, Giménez-Arnau A,Godse K, Gonc¸ alo M, Grattan C, Hebert J, Hide M, Kaplan A,Kapp A, Abdul Latiff AH, Mathelier-Fusade P, Metz M, Nast A,Saini SS, Sánchez-Borges M, Schmid-Grendelmeier P, Simons FE,Staubach P, Sussman G, Toubi E, Vena GA, Wedi B, Zhu XJ,Maurer M, European Academy of Allergy and Clinical Immunology, Global Allergy and Asthma European Network, European Derma-otology Forum, World Allergy Organization (2014) The EAACI/GALEN/EDF/WHO guidelines for the definition, classification, diagnosis and management of urticaria. Allergy 69(7):868–887