MAPÚA MALAYAN COLLEGES MINDANAO

Skin Disease Identification through Image Classification and Segmentation Using Deep Learning Techniques

Emmanuella Jester R. Casis
Andre Miguel C. Bacaling
Joan S. Gumban

Bachelor of Science in Computer Science

Thesis Adviser

Genevieve A. Pilongo, DBM - IS
College of Computer and Information Science

Skin Disease Identification through Image Classification and Segmentation Using Deep Learning Techniques

by

Emmanuella Jester R. Casis
Andre Miguel C. Bacaling
Joan S. Gumban

A Thesis Report Submitted to the College of Computer and Information Science
in Partial Fulfillment of the Requirements for the Degree

Bachelor of Science in Computer Science

Mapúa Malayan Colleges Mindanao
Davao City

May 2023

APPROVAL SHEET

The thesis, entitled **"Skin Disease Identification through Image Classification and Segmentation Using Deep Learning Techniques"** prepared and submitted by **Andre Miguel C. Bacaling, Emmanuella Jester R. Casis, Joan S. Gumban** in partial fulfillment of the requirements for the degree of **Bachelor of Science in Computer Science** is hereby accepted.

Genevieve A. Pilongo, DBM - IS

(Signature over printed name)

DEAN

Accepted as partial fulfillment of the requirements for the degree of **BACHELOR OF SCIENCE IN COMPUTER SCIENCE**

Rhodessa J. Cascaro, DIT

(Signature over printed name)

DEAN

ACKNOWLEDGEMENT

First, the proponents would like to express our deepest gratitude to the institution of Mapúa Malayan Colleges Mindanao, for allowing us to conduct this study within our community. The school has provided us with a world-class education and without that, we would not have been able to conduct this research.

We would also like to express our deepest appreciation to our Research Department Head, Dr. Gernelyn Logrosa. Her knowledge, dedication, and leadership skills have been pivotal in achieving the department's goals and objectives. Her guidance and support towards us seniors allowed us to produce high-quality research that has made a significant impact on our field.

To our dear adviser, Ma'am Genevieve Pilongo, we cannot thank you enough for your guidance in writing and developing our study. We are sincerely thankful for your support, motivation, and your willingness to help with our research paper. Thank you for believing in our capabilities to do this study. Without your help, this paper would not have been as feasible and comprehensive.

We also would like to express our deepest thanks to our panelists, Sir Martzel Baste, Sir Patrick Cerna, and Sir Christopher Josh Dellosa, for providing us with valuable feedback and suggestions during our thesis defense.

Finally, we are truly grateful to our families and friends for their never-ending support, love, and understanding throughout the time of conducting this study.

To God be all the glory and honor!

TABLE OF CONTENTS

LIST OF TABLES

ARTICLE 1

LIST OF FIGURES

ARTICLE 1

LIST OF ABBREVIATIONS

CNN	Convolutional Neural Network
DOH	Department of Health
HAM10000	Human Against Machine with 10000 training images
HFMD	Hand, Foot, and Mouth Disease
HPV	Human Papillomavirus
ISIC	International Skin Imaging Collaboration
KNN	K-Nearest Neighbors
LSTM	Long Short-Term Memory
ML	Machine Learning
ReLU	Rectified Linear Unit
RESNET	Residual Network
SVM	Support Vector Machine
UI	User Interface
VGG	Visual Geometry Group

Article 1

Skin Disease Identification through Image Classification and Segmentation Using Deep Learning Techniques

Andre Miguel C. Bacaling[a*], Emmanuella Jester R.Casis[a], Joan S. Gumban[a]

[a]College of Computer and Information Science, Mapúa Malayan Colleges Mindanao, Davao City, Philippines 8000

Emails: amBacaling@mcm.edu.ph, ejCasis@mcm.edu.ph, jGumban@mcm.edu.ph

*Corresponding author

College of Computer and Information Science

Mapúa Malayan Colleges Mindanao

Gen. Douglas MacArthur Hwy, Talomo

Davao City, Philippines

Email amBacaling@mcm.edu.ph

Abstract

The biggest organ in the human body is the skin, and just like any other organ, the skin can also have numerous problems such as infections which are commonly diagnosed through traditional methods that include physical examinations in the affected area, studying the patient's medical history, etc. Such methods can be time-consuming, costly, inaccurate, and unavailable in other areas of the world. To address this, this study entitled, "Skin Disease Identification through Image Classification and Segmentation Using Deep Learning Techniques" was conducted. The focus of this was the most common viral skin infections, such as chickenpox, monkeypox, warts, corns, eczema, and warts. The approach of this research was designed to provide a method of detecting these skin infections that is efficient, accurate, and low to no cost. In support of its goal to detect such skin diseases, a predictive model, a web-based application to visualize and test the model through image classification and segmentation using deep learning was developed. The dataset consists of 700 images collected from sources such as World Health Organization (WHO), Kaggle, DermNet, and various dermatological websites such as WebMD and MayoClinic, which are dedicated to skin diseases. The data were divided into two parts, the training augmented dataset and validation data. The augmented data was used to train the model while some of the validation data was used to determine whether the model was functioning properly. Based on the testing and results, the model developed was able to predict corn with 75% accuracy, Chickenpox with 99% accuracy, Eczema with 81% accuracy, and Monkeypox with 99% accuracy. Warts were predicted with 89% accuracy.

Keywords: Skin Disease, Image Classification, Segmentation, Deep Learning Technique

1. Introduction

The biggest organ in the human body is the skin, and just like any other organ, the skin can also have problems, such as cradle cap rashes from newborn babies to age spots in elders. There are many different categories of skin problems. There are skin discolorations, chronic skin conditions, skin cancers, acute skin problems, and viral skin infections (“A Guide” 2021). This study focused more on viral skin infections.

Globally, millions of individuals are afflicted by the widespread and incapacitating group of viral skin infections (Johnson, 2023). These infections can appear in many different forms, including warts, cold sores, monkeypox, shingles, chickenpox, etc. Viral skin infections arise because they are manifested by pathogens such as the Human Papillomavirus (HPV), Herpes simplex virus, and varicella-zoster virus. Skin disease and infections are more prevalent in developing countries and tropical climates. In the Philippines, viral skin infections are also prevalent. The most common infections that can be found in the Philippines include warts, eczema, and psoriasis (De Goma & Devaraj, 2021). Locally, in Davao City, there were news reports on skin disease outbreaks such as Hand, Foot, and Mouth Disease (HFMD), and measles (Mendoza, 2023). These skin infections have different traits of what they can do to the human body. It can be a mild irritation or even a severe condition. Some are even highly contagious and could be at risk of an outbreak.

With the presence of such diseases which may potentially threat people’s health, this study was conducted with the aim to provide a way to identify these skin diseases early on through the help of technology and the use of Machine Learning to avoid or prevent an outbreak from happening. In recent years, Machine Learning has become a powerful tool for image classification and analysis (Sarker, 2021). It is widely used to solve real-world problems, including in the medical field. This study utilized deep learning techniques, a subset of Machine Learning, to detect viral skin lesions through image processing and classification, an accurate, and efficient tool for healthcare facilities and patients.

There are traditional methods for identifying viral skin infections. These include physical examinations in the affected area, studying the patient's medical history, and performing lab tests such as Polymerase Chain Reaction (PCR) tests, a laboratory technique used to detect and amplify specific segments of DNA or RNA in a sample, and serological tests. There is also a method wherein it involves scraping off a small part of the affected skin and then examining it under a microscope. These traditional methods for identifying skin infections can be time-consuming, costly, inaccurate, and unavailable in other areas of the world. This study provides a way of detecting these skin infections efficiently, accurately, and with little to no cost using Deep Learning.

Deep learning is a subset of machine learning. Relatively, it is a new technology that has been proposed as a tool for diagnosis and classifying various medical conditions such as skin diseases. In recent years, research papers such as "Early Detection of Skin Cancer - Solution for Identifying and Defining Skin Cancers using AI" (N. Darapaneni et al., 2022) and "Improvement of Skin Cancer Detection Performance Using Deep Learning Technique" (F. Yilmaz et al., 2020) are some of the several studies that have used deep learning as a means of diagnosing skin cancers. In the case of viral skin infections, its application of deep learning is a relatively recent area of study. However, there is already some research about the application of deep learning for identifying viral skin infections such as those conducted by Yu-Zhu et al. (2021) and Alghieth (2022) which focused on diagnosing skin diseases in clinical environment and in schools, respectively.

In these studies, a deep learning model was often trained on a dataset of skin disease images, its diagnostic performance was compared to that of human experts. Convolutional Neural Networks (CNNs), which are deep learning architectures particularly well adapted for image classification problems, are frequently used in investigations. The main goals of their studies were to evaluate the diagnostic precision and accuracy of deep learning models for viral skin infections and to ascertain if they can be applied as diagnostic tools in clinical settings. Although deep learning has demonstrated promise in diagnosing skin conditions, such as skin cancer detection using deep learning techniques by Dildar et al. (2021), more research is needed to determine how well it performs in detecting viral skin infections. It is also crucial to note that before these models for identifying skin infections can be utilized as a diagnostic tool, their accuracy must be confirmed in bigger and more varied patient populations as well as in real-world clinical settings.

Table 1. ML algorithms used by previous research

No.	Research Title	EfficientNet-b4	Inception-v3	ResNet-50	VGG-19	AlexNet	SVM	Others (KNN, DT, etc.)
1.	A Deep Learning Based Framework for Diagnosing Multiple Skin Diseases in a Clinical Environment	✓	✓					✓
2.	A machine learning approach for skin disease detection and classification using image segmentation					✓	✓	
3.	Skin Disease Detection for Kids at School Using Deep Learning Techniques				✓			
4.	Skin Disease Detection Using Deep Learning			✓			✓	✓

Table 1 shows that each research article used different types of pre-trained models. Some also created their model or modified the pre-trained models to achieve better results. In this study, the researchers used the ResNet-152 pre-trained model as only a handful of research about skin disease detection have had it and it is also one of the models that provide high accuracy in image classification.

Table 2. Types of skin problems included in previous research

		Skin Problems Used					
No.	Research Title	Monkeypox	Chickenpox	Corns	Eczema	Warts	Others
1	A Deep Learning Based Framework for Diagnosing Multiple Skin Diseases in a Clinical Environment				✓	✓	✓
2	A machine learning approach for skin disease detection and classification using image segmentation						✓
3	Skin Disease Detection for Kids at School Using Deep Learning Techniques		✓			✓	✓
4	Skin Disease Detection Using Deep Learning				✓		✓

Table 2 shows the differences between the previous related research articles based on the skin problems they have included as their dataset. It shows the different types of skin problems that are included in the said research articles. Moreover, it was noted these articles focused more on other skin problems, specifically the different types of skin cancers. In this study, the research only concentrated on viral skin infections, a different category of skin problems.

The general purpose of this study was to detect some of the most common skin diseases using deep learning techniques and by developing an improved or modified predictive model from previous related studies.

Specifically, it sought to answer the following questions:

1. What deep learning techniques can best apply in image classification and segmentation to identify skin diseases?
2. What deep learning platform is best used for implementing these deep learning techniques?
3. What is the accuracy of the algorithm developed in terms of identifying a skin disease?

The study objective was to develop a web-based application to visualize and test the model through image classification and segmentation using deep learning. Specifically, this study aims to:

1. Develop a responsive web-based application with the integration of a pre-trained model made by a convolutional neural network architecture algorithm.

2. Utilize the FastAI deep learning library to construct a convolutional neural network model for image classification, using an augmented training dataset of over 3,000 images. Eighty percent (80%) of the dataset would be used for validation and the remaining 20% would be used for training, with an overall accuracy goal of 90%.

3. Determine the accuracy of the predictive model by comparing its results with actual data.

This study was undertaken to develop a soft web application that could analyze the uploaded image of the user and classify whether the patient potentially has a viral skin infection or not. This study may be beneficial to the following:

Medical Team. This study may help medical teams to classify whether the patient has the virus.

Other Researchers. This study will add to the body of knowledge that other researchers may use as a reference for a similar study.

2. Related Works

2.1 Skin Diseases

Skin disease is just one of the many diseases in the world that affects millions of people worldwide. In this study, the researchers will create an application to detect skin diseases particularly monkeypox, chickenpox, warts, eczema, and corn.

According to the World Health Organization, monkeypox occurs in central and west Africa and has increasingly appeared in urban areas. The first case of monkeypox was identified in humans in 1970 in Congo. The first case identified outside Africa was in the United States of America and it was linked to contact with infected pet prairie dogs.

Another viral skin disease is chickenpox. Chickenpox is caused by the varicella-zoster virus that belongs to the a-herpes family. The virus is highly infectious, and it is contracted through person-to-person contact. Even though chickenpox and monkeypox both have “pox” in their names, the said diseases are different and caused by unrelated viruses. Chickenpox is caused by the varicella-zoster virus while monkeypox is caused by the monkeypox virus, a member of the orthopoxvirus genus in the family Poxiridae.

The third skin disease is called warts. Warts are caused by HPV or known as human papillomavirus. The virus is contagious, and people can get warts from touching someone who was them. It commonly appears on the hands, but it can also appear on the feet, face, genitals, and knees. Most warts go away after the immune system fights off the virus, but

some warts cause issues like cancer, disfigurement, infection, and pain. Genital warts are linked to different cancers like cervical cancer and anal cancer (Cleveland Clinic)

The fourth skin disease is called eczema. It is a condition where the skin becomes dry, itchy, and bumpy. According to National Eczema Association, more than 31 million Americans have eczema. It can begin during childhood, and it can range from mild to severe. There is no cure for eczema but there are treatments that can help manage and minimize the symptoms.

The fifth skin disease is Corns. It is a thick, hardened layer of skin that develops when the skin protects itself from pressure. Some people distinguish Corns and calluses as the same, but they are different. Corns are smaller and deeper than calluses and they are also painful when pressed (Mayo Clinic).

According to Skin MD, skin diseases have low mortality rates, but they can be an indicator of underlying medical conditions and may also lead to health complications if not treated. The most common skin diseases in the Philippines are warts, eczema, hair, and nail problems.

2.2 Skin Diseases Using Image Classification

Skin disease detection is being studied by many researchers in line with machine learning, deep learning models, and image processing.

Many researchers made skin disease detection studies with different algorithms and deep learning models. A study by Algeith (2022) entitled “Skin Disease Detection for Kids at School Using Deep Learning Techniques” used the Convolutional Neural Network (CNN) technique and a pre-trained Visual Geometry Group 19 (VGG19) model to detect skin disease. VGG19 model is a CNN that is 19 layers deep. It is the most used image-recognition architecture, and it proves to have a high accuracy result. The study by Algeith achieved a high accuracy of 99%.

Kshirsagar et al. (2022) applied MobileNetV2 and Long Short-Term Memory (LSTM) Networks to develop a skin disease classification system. Based on the result of the project, the model surpassed other techniques and models to detect skin diseases.

A study by Al Shabibi and Koottala (2020) used Matlab to detect acne, cancer, and psoriasis dermatitis. They also utilized SVM to identify the symptoms of the three skin diseases to improve the identification accuracy.

The research by Ahammed et. al. (2022) entitled “A machine learning approach for skin disease detection and classification using image segmentation” introduced a digital hair removal technique. They also applied the automatic Grabcut segmentation technique to segment the affected lesions. Three machine learning techniques, Decision Tree, Support Vector Machines (SVM), and K-Nearest Neighbors (KNN) were used to classify the skin images. The models of this study were validated using two datasets, International Skin Imagine Collaboration (ISIC) 2019 challenge and Human Against Machine with 10000 training images (HAM10000). The result shows that SVM performs better than the two classifiers.In this study by Codella et. al. (2018) entitled “Skin Lesion Analysis Towards Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI)”, the authors used the Inception v3 model as a feature extractor for images of skin lesions and then trained a simple linear classifier on top of the extracted features to classify the images as either benign or malignant. The study participated in the ISBI 2017 challenge on skin lesion analysis and achieved state-of-the-art performance, outperforming other methods that used more complex architectures or additional data. The study showed that using a pre-trained model such as Inception v3 can be a good choice for skin lesion classification tasks, as the model has already learned features that are useful for image classification in general.

In this study, "Skin Lesion Analysis Towards Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI)," the authors used a DenseNet-121 architecture and trained it on the ISIC 2017 Challenge dataset. They achieved an overall accuracy of 87.4% on the test set, which was among the top results in the competition. Another study is "Densely Connected Convolutional Networks for Skin Lesion Analysis" in which they used DenseNet-169 architecture and achieved an accuracy of 89.2%. These studies demonstrated the potential of using DenseNet for skin lesion classification.

In the study of De Goma, J.C., and Devaraj, M. (2020b) entitled Recognizing Common Skin Diseases in the Philippines Using Image Processing and Machine Learning Classification, the proponents created a system that detects acne vulgaris, atopic dermatitis, keratosis pilaris, psoriasis, leprosy, and warts using different pre-processing and segmentation algorithms. The proponents used the SVM classifier and Artificial Neural Network (ANN) classifier to train the model. For the SVM classifier, the model accomplished an average of 93.55% and 93.33% for precision and recall and for the ANN classifier, it accomplished an average of 96.55% and 100% for precision and recall.

3. Methods

3.1 Research Participants

The participants in this research consisted of doctors and physicians who evaluated the output of this project. Their role was to test the application's validity and assess its performance.

3.2 Research Locale

The research was conducted in Davao City, Philippines. The researchers implemented the methodology within the convenience of their own homes as well as within the premises of the school.

3.3 Project Procedure

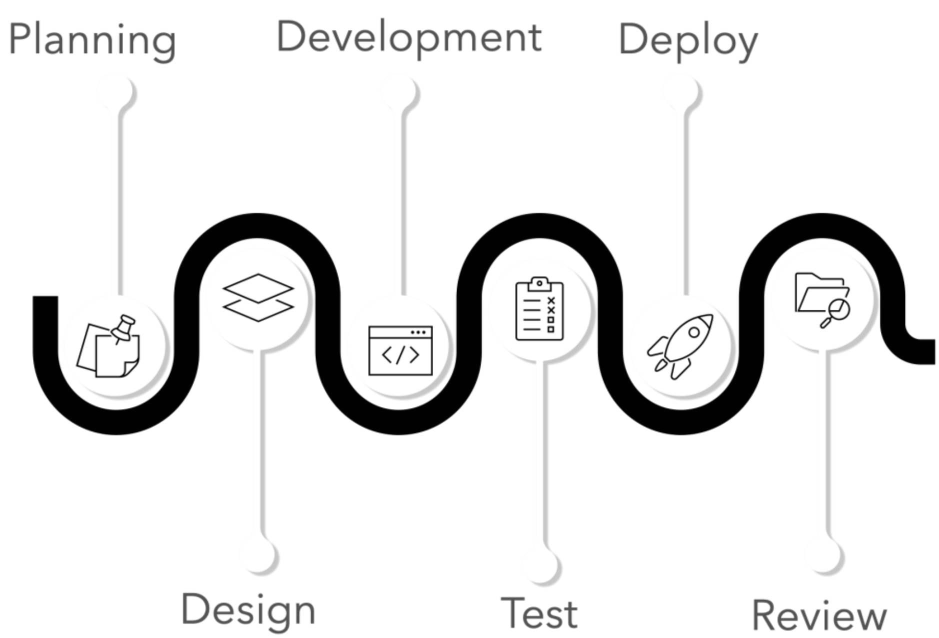

Fig. 1. AGILE software development process

Planning

The planning phase is the phase wherein the researchers discussed the approach of the research, such as what tools were needed to develop the application, and where to find reliable image datasets for the model to learn.

Design

In this phase, the researchers designed a prototype of their website, what the app should look like, and what functionalities it should have. This is also the phase where the researchers made the initial manuscript.

Development

In this phase, this is where the design was put into action. The researchers started to develop the website. The researchers started to make the front end and develop the back end.

Test

The testing phase is where the model training and model evaluation took place. In this phase, the researchers tested the system to check if there were any bugs or errors in the application.

Deploy

The deployment phase is where the developed system is ready to be deployed and hosted to a web server for the public to use. In this phase, the application was ready for deployment.

Review

After the system is deployed, this is the phase where a reviewing phase was initiated. The researchers conducted an evaluation of the performance of the application, how well it performed, and if there was a need to have changes.

3.4 Data Collection

Table 3. Skin Disease Dataset

Disease	Sample Images	Number of Images
Monkeypox		279
Corns		36

Table 3 (cont'd)

Eczema		180
Chickenpox		107
Warts		101

The dataset consists of 703 images collected from sources such as WHO, Kaggle, DermNet, and various websites such as WebMD, Clevelandclinic, and MayoClinic that also discuss skin diseases. The data have been divided into two parts the training augmented dataset and validation data. The augmented dataset was utilized for training the model, while a portion of the validation data was used to evaluate the model efficiency. The collected data was categorized by the researchers into distinct sections, representing various types of viral diseases. The researchers have included viral diseases, which have been prevalent worldwide.

Table 4. Data Source

Data Source
World Health Organization
Kaggle
DermNet
WebMD, Clevelandclinic, MayoClinic

Table 4 presents the various sources from which the data was collected. The World Health Organization (WHO) is a global agency focused on promoting health and ensuring global safety. Kaggle is a website that provides datasets for users conducting studies. DermNet is a website dedicated to dermatology and skin conditions. WebMD is an online publisher that focuses on human health and well-being. The Cleveland Clinic is a nonprofit medical center renowned for its clinical and hospital care. Lastly, the Mayo Clinic is another nonprofit organization dedicated to clinical practice and research, providing comprehensive care for individuals in need of healing.

3.5 Image Augmentation

Image augmentation is a technique that is used in computer vision to artificially increase the size of a training dataset by generating variations of existing images. The purpose of this technique is to enhance the diversity of the images used to train models in machine learning, which can lead to a better generalization on new, and unseen data. Image augmentation involves a set of transformations to the original images, such as flipping, rotating, cropping, zooming, adding noise or distortions, adjusting brightness and contrast, and changing the balance of color and hue. These transformations can be applied randomly or systematically to generate multiple variations of each image, thereby creating an augmented dataset. In this study, the researchers employed data augmentation techniques on the acquired dataset to enhance the model's accuracy.

Table 5. Total Amount of Data Before and After Augmentation

Disease	Amount of Data Before Augmentation	Amount of Data Before Augmentation
Monkeypox	279	1,116
Corns	36	360
Eczema	180	720
Chickenpox	107	428
Warts	101	404
Total Number of Data	703	3,028

Table 5 presents the comparison between the amount of data on each skin disease and its total amount before and after augmentation. Before augmentation, monkeypox had

279 images, corns had 36 images, eczema had 180 images, chickenpox had 107 images, and warts had 101 images with a total amount of 703 images. After augmentation, monkeypox had 1,116 images, corn had 360 images, eczema had 720 images, chickenpox had 428 images, and warts had 404 images with a total amount of 3,028 images.

3.6 Preprocessing

Fig. 2 Process flow classification algorithm Resnet

Figure 2 reveals the process flow of the classification algorithm using Resnet. The input skin images used as input were obtained from patients diagnosed with the disease during the study. As part of the pre-processing step, the images were resized to ensure accurate dimensions and a Gaussian noise function was applied. To extract features from the images, the researchers implemented an algorithm that operated on multiple layers of the image. This requirement led the researchers to use the Resnet 152 layers algorithm.

To achieve accurate detection and prediction of skin diseases, the researchers encountered several challenges. Initially, training the model using the original dataset yielded poor results. One of the main issues Identified was the imbalance of the dataset, where certain classes had significantly more samples than others. To address this problem, the researchers implemented a solution by down sampling the dataset to approximately 1000 samples per class. Additionally, they introduced Gaussian noise to each image as part of the data augmentation process. These measures were taken to mitigate the impact of data imbalance and improve the overall accuracy of skin disease detection and prediction. The process involved discarding alternate pixels in both the horizontal and vertical dimensions of the images. Subsequently, the modified images were converted into comma-separated value (csv) format to use as a basis for the augmented dataset. Gaussian noise could be added to images to create new images that look similar but not identical. This technique served two purposes: it helped to prevent overfitting of the machine learning algorithm. Additionally, it can increase the adaptability of the algorithm to noise and other variations in input data.

To extract features from an image, the researchers required an algorithm that can operate on different layers of the image. To achieve this goal, Randoms liter was used for splitting the training and validation dataset and performed item level transforms on input data to resize image to 224 pixels. Additionally, batch level transforms to the input data I-E. A random resized crop of size 224 pixel was applied using GPU. The detection of skin depends upon the gaussian noise pixel color ration of each pixel. The detection of skin depends upon pixels and Gaussian noise of each pixel. The XY values can be generated from the ration of gaussian noise for each pixel and input into a formula. A (x, y) = H (x, y) + B (x, y) Where, A (x, y)= function of noisy image. H (x, y) = function of image noise, B (x, y) = function of original image.

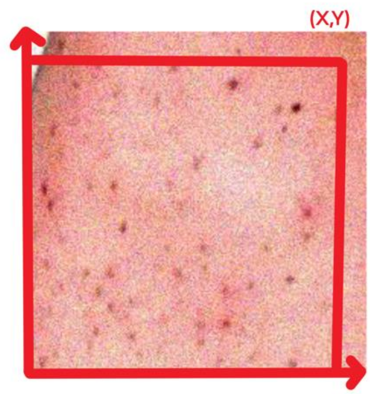

Fig. 3. Gaussian Noise applied in an Image.

The data I-E I (x, y) is an image function where I = Intensity at position (X,Y) in an image. The researchers created data loaders for the training dataset to label groups of pixels of each image. Identification of the infected area helped the developed model to ease the prediction of the classifier to find the disease.

3.7 Feature Extraction

	image_id	label	label_name
0	chickenpox/chickenpox95_270.jpg	1	chickenpox
1	chickenpox/chickenpox95.png	1	chickenpox
2	chickenpox/chickenpox95_90.jpg	1	chickenpox
3	chickenpox/chickenpox95_mir.jpg	1	chickenpox
4	chickenpox/chickenpox75_90.jpg	1	chickenpox
...	...	...	...
5375	warts/warts-plantar-8_mir.jpg	6	warts
5376	warts/warts-flat-21.jpg	6	warts
5377	warts/warts-flat-21_90.jpg	6	warts
5378	warts/warts-flat-21_mir.jpg	6	warts
5379	warts/warts-flat-21_270.jpg	6	warts

5380 rows × 3 columns

Fig. 4. Feature Extraction of the Dataset

Figure 4 shows the CSV file to store information about image datasets. This involved creating two separate files: one for the dataset itself and another for labels. The dataset CSV file lists all the images in the skin diseases dataset along with their file paths and any additional relevant metadata. The labels CSV file contains a list of all the images in the dataset along with their corresponding labels. Overall, Csv files can be used for data visualization and analysis of image in data, making it easier to understand the dataset and derive insight from it (“Data Analysis”, 2023).

The researchers created a Datablock object from a CSV file, they used from_csv() method provided by the fastai library. This method takes the path of the CSV file and the names of the input and target label columns as input. After reading the CSV file, it splits the data into training and validation sets using a splitter function and applies any necessary preprocessing steps to the data. The Random Splitter function is used as a splitter function to split the data into a training dataset and a validation dataset with a split ratio of 80:20 (80% training and 20% validation). Finally, the Datablock object obtained can be used to train a machine learning model using the fastai learner.

3.8 Classification

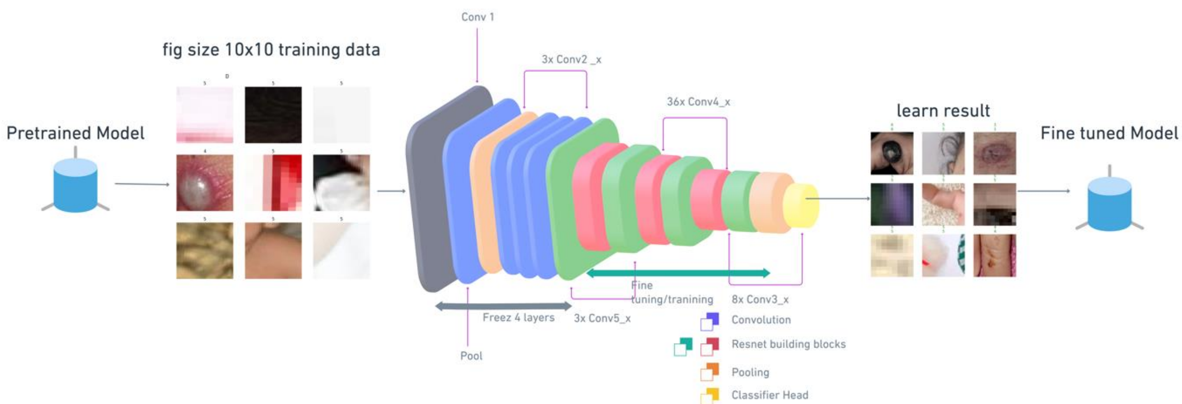

Fig. 5. Architecture of ResNet152

Figure 5 shows the Architecture of ResNet-152 in this study. ResNet-152 is a type of neural network architecture used for image classification containing 152 layers. This architecture consists of an input layer with RGB channels of size 224x224x3, followed by several convolutional layers with varying filter sizes and channel numbers. There were four stages, each with a stride of 2, that down-sample feature maps. Within each stage, there were residual blocks that contained two 3x3 convolutional layers, a shortcut connection, and batch normalization with Rectified Linear Unit (ReLU) activation function. The feature maps were then averaged spatially through global average pooling to produce a

2048-dimensional feature vector. This vector was then fed into two fully connected layers, which are followed by a softmax layer that outputs the predicted probabilities for each class. ResNet-152 is a sophisticated neural network architecture with many parameters. Nevertheless, its residual learning design allows for efficient training and better performance in image classification tasks.

3.9 Resources and Model Description

FastAI

Fastai is a library for deep learning that offers pre-built tools for practitioners to achieve high-quality results in common deep learning tasks with ease. It also provided the researchers with customizable building blocks to develop new approaches without sacrificing usability, adaptability, or efficiency. This is through a layered architecture that uses abstract concepts to represent the fundamental patterns of various deep learning and data processing techniques. These abstractions are expressed clearly and briefly by utilizing the flexibility of the PyTorch library and the dynamic nature of the Python language. Fastai has its own advantages compared to other deep learning libraries and we can use it to build and train image processing on the model.

Fine-tuning

Fine-tuning is to leverage the knowledge learned from a pre-trained model and apply it to a new task. Fine-tuning can be done by adjusting the parameters of the pre-trained model or by training the entire model on the new dataset. It is a popular technique used to improve the performance of deep learning models, especially when the new dataset is small or like the original dataset used to train the model.

Flask

Flask is a micro web framework for python that allows developers to build web applications quickly and easily. The front-end of a Flask application refers to the part of the application that interacts with the user, such `as the HTML, CSS, and JavaScript that make up the user interface.

OpenCV (Computer Vision)

OpenCV is a powerful library for image processing, which is one of its primary focus areas. It provides a wide range of functions for image manipulation, including image filtering, thresholding, color space conversions, feature detection and extraction, and image alignment and registration.

3.10 Project UI

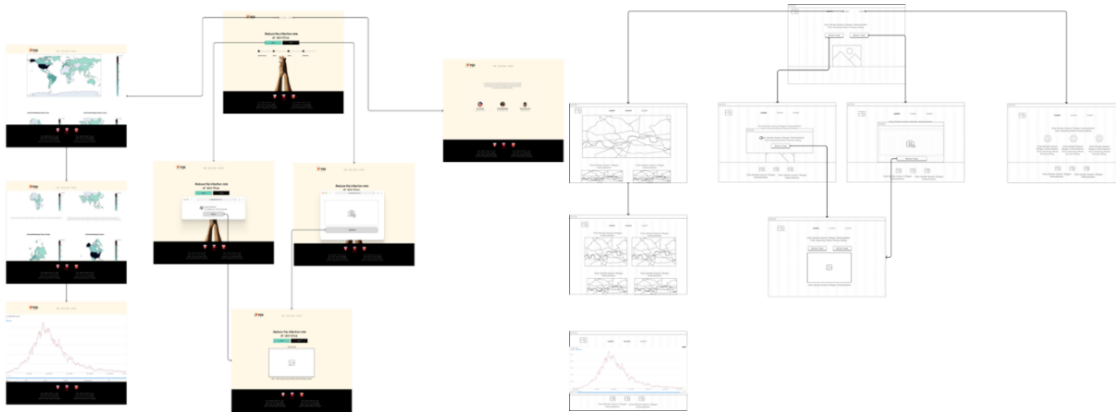

Fig. 6. Project User Interface

Figure 6 shows the system's User Interface (UI). It displays all the pages that are in the system.

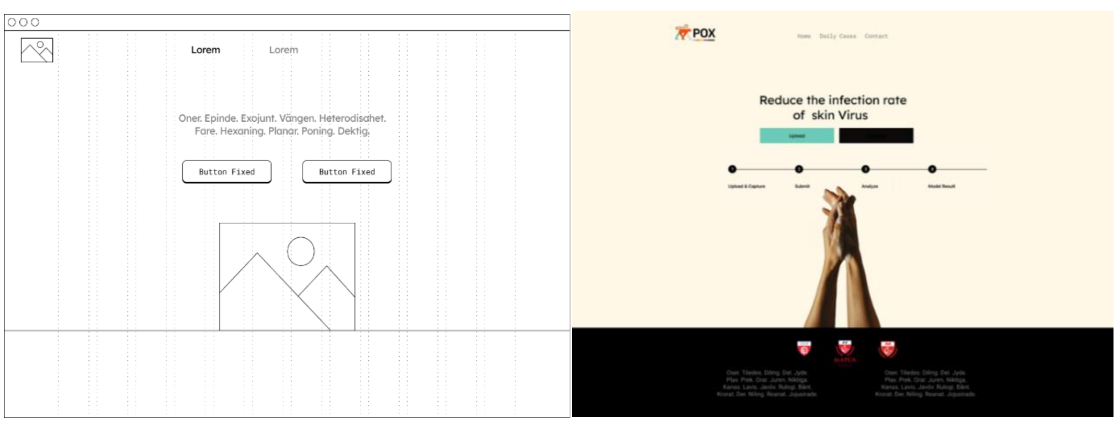

Fig. 7. Home Screen

Figure 7 shows the home screen of the disease's detection. This screen appears when the application's URL is entered, and it includes a system guideline for the user to view.

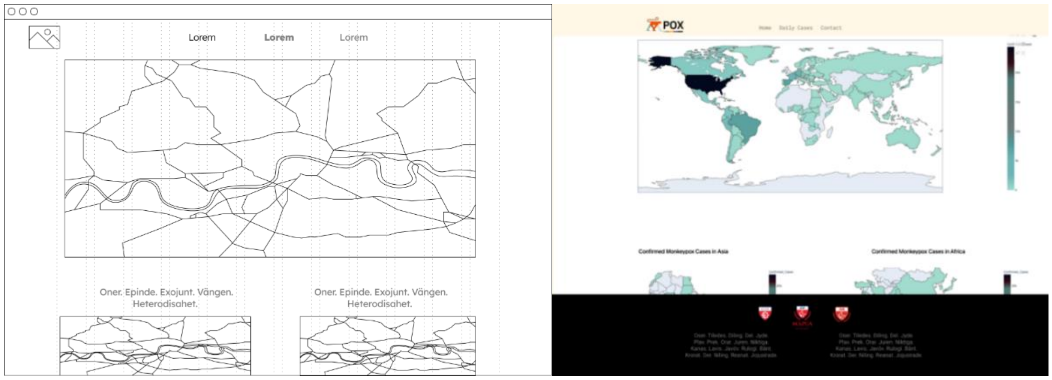

Fig. 8. Analysis Cases Screen

Figure 8 shows the analysis cases. The system has a feature that utilizes the Global Viral skin diseases dataset from Kaggle for the purpose of analyzing cases. These datasets are displayed using an open-source library called plotly.js, which can be visualized in Jupiter notebook and integrated into the server.

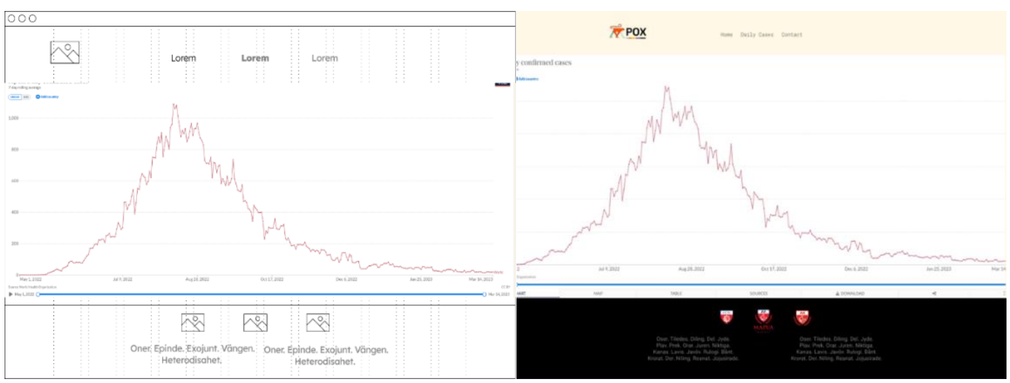

Fig. 9. World Health Organization Viral Skin Diseases Analysis Cases Screen

Figure 9 shows the WHO's Viral Skin Diseases Analysis Cases, an HTML iframe element tag that embeds and interactive chart or table from the website ourworldindata.org. The chart is related to viral disease cases today and displays confirmed cases data over time, with options to filter by various metrics such as frequency and relative population.

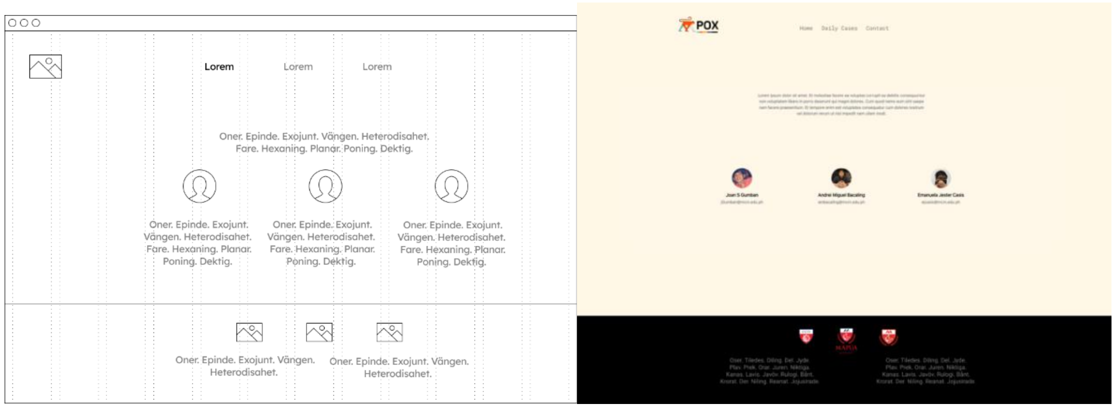

Fig. 10. Contact Screen

Figure 10 shows the Contact Screen. This feature allows the user to view the developers of the application, making it easier for the user to reach out to them.

3.11 Conceptual Framework

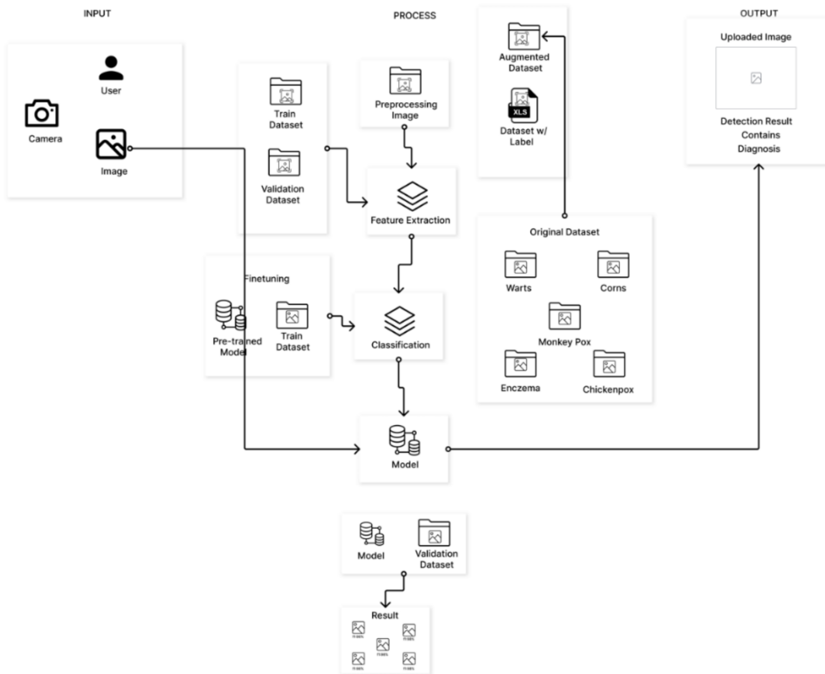

Fig. 11. Conceptual Framework

Figure 11 shows the architecture of the application. As the user launches the app, it starts with a home screen. Users can see the guidelines which contain the process of how to use the application. In the home screen, the user can pick between the upload image button or the capture button. Both options have the same function: to identify what diseases the user has; it also shows the accuracy of the prediction and the diagnoses.

3.12 System Flow

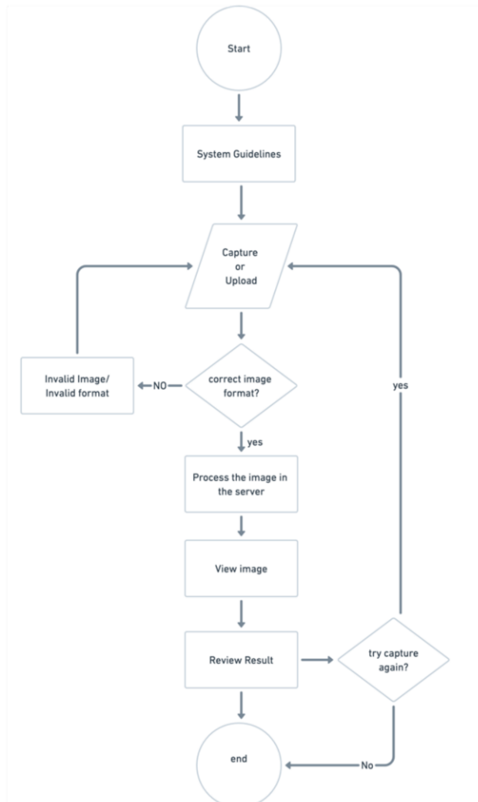

Fig. 12. System Flow

Figure 12 is the flowchart for the proposed system. It starts with the system guidelines, then it will lead you to an option to either upload an image from your device or capture an image using your device's camera. If the image is the correct image format, it will be fed to the system's server. Otherwise, it would not be accepted. After an image has been fed to the system, it will be sent to the server for it to be predicted and classified by the training model. After the process, the model's classification and prediction rate will be shown to the user.

The main flow of the system are as follows:

1. The user opens the web-app on their device.
2. The web-app prompts the user to grant access to the device's camera.
3. The user grants camera access.
4. The web-app displays a camera viewfinder on the screen.
5. The user positions the camera to take a photo of the skin lesion.
6. The user captures a photo of the skin lesion using the device camera.
7. The web-app displays a prompt indicating that the image is being analyzed.
8. The skin disease classifier analyzes the captured image and identifies the type of skin disease.
9. The web-app displays the predicted skin disease with the corresponding probability score.

An alternate flow of this use case diagram would be if the user's device does not have a camera, the web-app has the option to upload an image of the skin lesion from their device's photo library.

3.13 Website Map

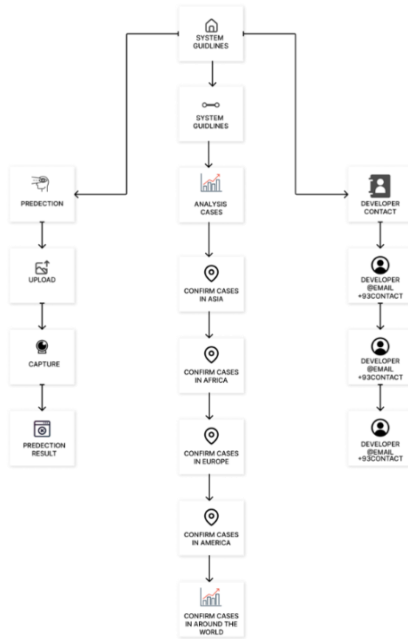

Fig. 13. Website Map

Figure 13 reveals the website map for the system. This website map guides the users to navigate through the different sections and pages of the website. When the user enters the website, they will be redirected to the home page and system guidelines. At the home page, the user will be met with three sections of the website: the prediction section, the real-time cases section, and the developer contact section. The prediction section is the main part of this project, and this is where the predictive model is used. This is also the section where the user will either upload an image of the unknown skin disease from their device or capture an image, to be predicted by the model. The real-time cases section will inform the user about daily cases of skin diseases around the world through data visualization. Lastly, the developer contact section is the section where the user will be able to retrieve the contact information of the developers or the researchers behind this project.

3.14 Data Augmentation Process

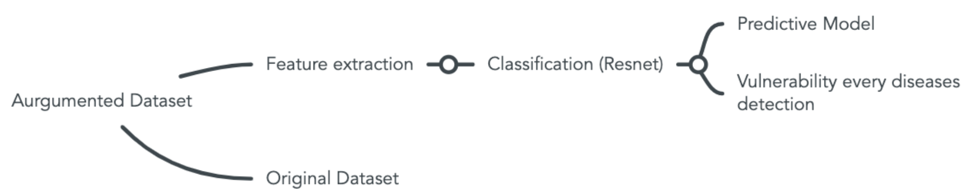

Fig. 14. Data Augmentation Process

Figure 14 shows the Data augmentation process. Data augmentation is a technique that is used to increase the number or size of the training dataset. This is used when there is a lack of number of images in a dataset. The goal of augmenting image data is to create additional images that are like that of the original image, but have different characteristics, such as the scale, cropping, angle, and lighting of the image. This process can help improve the performance of the machine learning model.

The first step in data augmentation is loading the original image dataset that is stored in a directory. The next step is to define a pipeline of transformation to apply to the original images. In this case, the researchers created a function to add gaussian noise to the image. Transformations have also been added such as flipping the image vertically and horizontally and rotating the image in a specific angle. Once a transformation pipeline has been defined, the images are now ready to be augmented.

3.15 Use Case

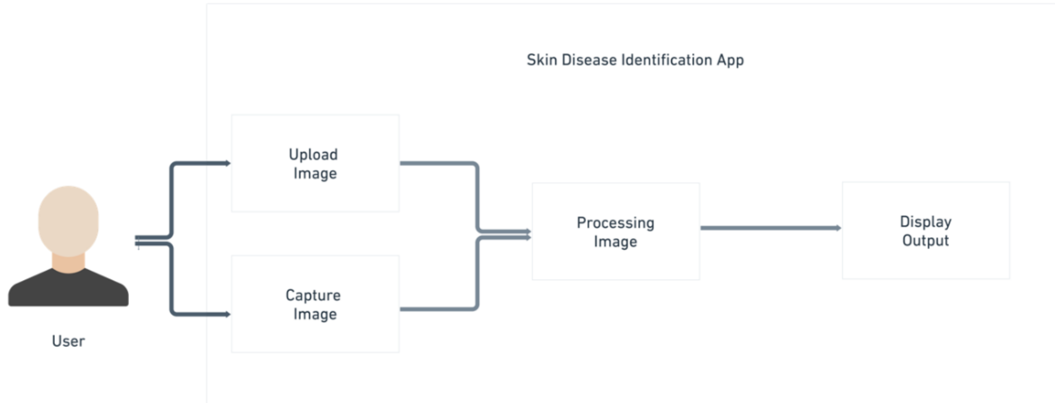

Fig. 15. Use Case Diagram

Figure 15 shows the Use Case Diagram of the researchers' skin disease identification web-application. The precondition is that the user must have access to the web-app and a device with a camera.

3.16 Limitations of Methodology

1. The researchers utilized data augmentation and down-sampling techniques to reduce the number of samples in each class to around 1000, and as a result, the researchers observed improvements in the f1 score.

2. Initially, the researchers tried training with the original dataset, but it resulted in a poor f1 score as the dataset was highly imbalanced. Therefore, the solution was to augment the images to attain a balanced f1 score.

3. The model's performance of the model was limited to the scope of the dataset it was trained on. This means that any other skin diseases, entities, and/or foreign objects that were not included or mentioned in the dataset were not predicted by the model.

4. Skin images that are captured or uploaded by the user to test the model should be clear and a close-up shot. The lighting and the ambience of the image is also important. It should be visible, not too bright, and not too dark. Otherwise, the prediction of the model will be affected.

3.17 Trustworthiness of the Study

The Reset and pre-trained models were used to help identify different kinds of skin lesions. These pre-trained models are commonly used in image classification. The researchers were confident that these models resulted in the intended outcome of this study.

3.18 Ethical Considerations

· Social Value

Identifying and treating skin diseases has social value by improving the quality of life for individuals, preventing outbreaks, reducing healthcare costs, and addressing health disparities. ·

·Informed Consent

By participating or testing out the system, the users acknowledged that they have received and understood the information provided, and consent that the image that they uploaded was used only for the prediction of the type of skin disease in the image.

· Privacy and Confidentiality

The datasets that were used in this research are strictly confidential. They were only used to fulfill the purpose of this research.

· Adequacy of Facilities

This study was conducted on devices that were always available to the researchers.

· Qualification of Researchers

The researchers are Computer Science students in Mapua Malayan Colleges Mindanao. They are qualified to conduct this study because they want to help the community in dealing with viral skin infections.

4. Results and Discussion

In this chapter, the results are presented and discussed. It consists of three sections, and the results of these sections are presented in accordance with the order of the three objectives of this study. Section 4.1 of the results is for objective 1; 4.2 for objective 2 and so forth.

4.1 Detection Analysis

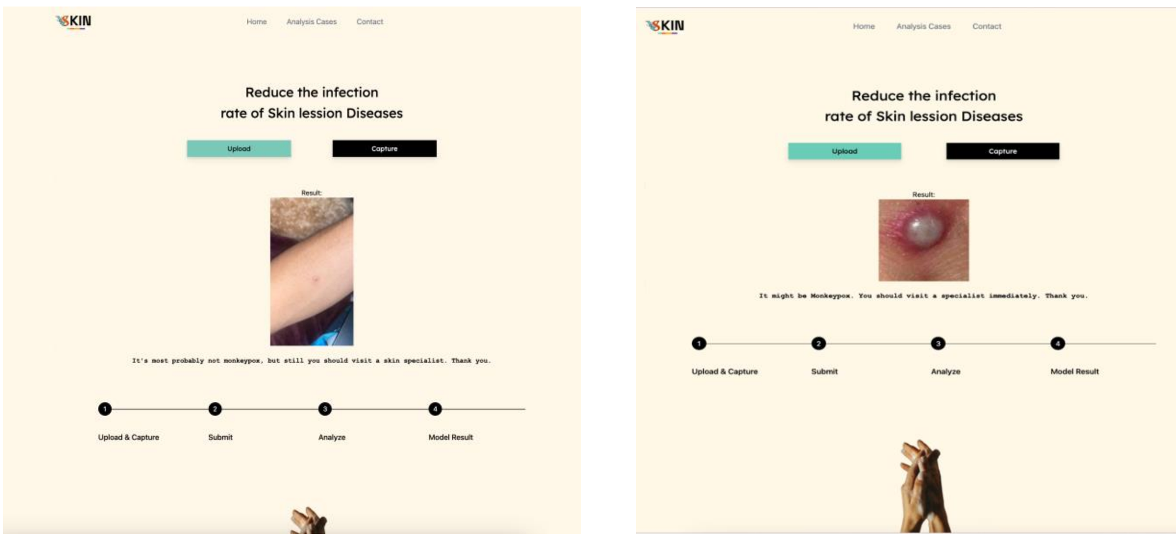

Fig. 16. Detection Analysis of Chickenpox, Corns, Eczema, Monkeypox, Warts

The Convolutional Neural Network was fed with trained data which helped in identifying the diseases. Features were extracted from the training dataset. A prebuilt model was used here to identify every layer's pixel of the disease. In classification, the researchers used label smoothing techniques for classification problems to prevent the model from predicting the labels too confidently during training. Meanwhile, for optimization, optimization function ranger was utilized.

Initially, to optimize the skin image and the model training result, the researchers applied some image processing methods by increasing the dataset using a function that generates 10 new images for each input image, including the original image and image with various rotations, mirroring, and Gaussian noise. This technique can be useful for increasing the size of the data and improving the training result of the machine learning model.

4.2.1 Data Distribution

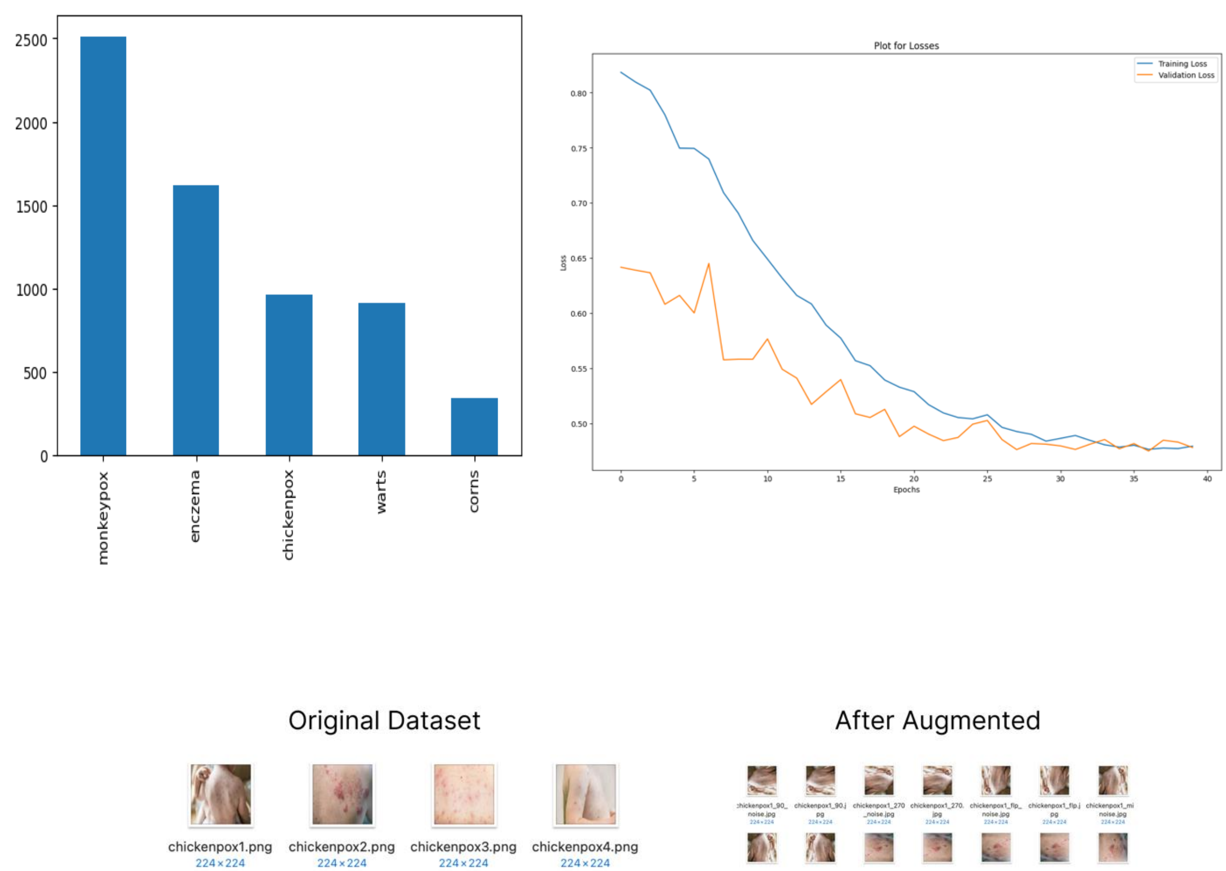

Fig. 17. Image Processing Optimization and Feature Extraction

Figure 17 shows the data distribution after the images have been augmented. Monkeypox images had the greatest number of augmented images as it was the dataset with the greatest number of original images. Meanwhile, corns had the least number of augmented images because of the lack of original images.

4.2.2 Model Classification Result

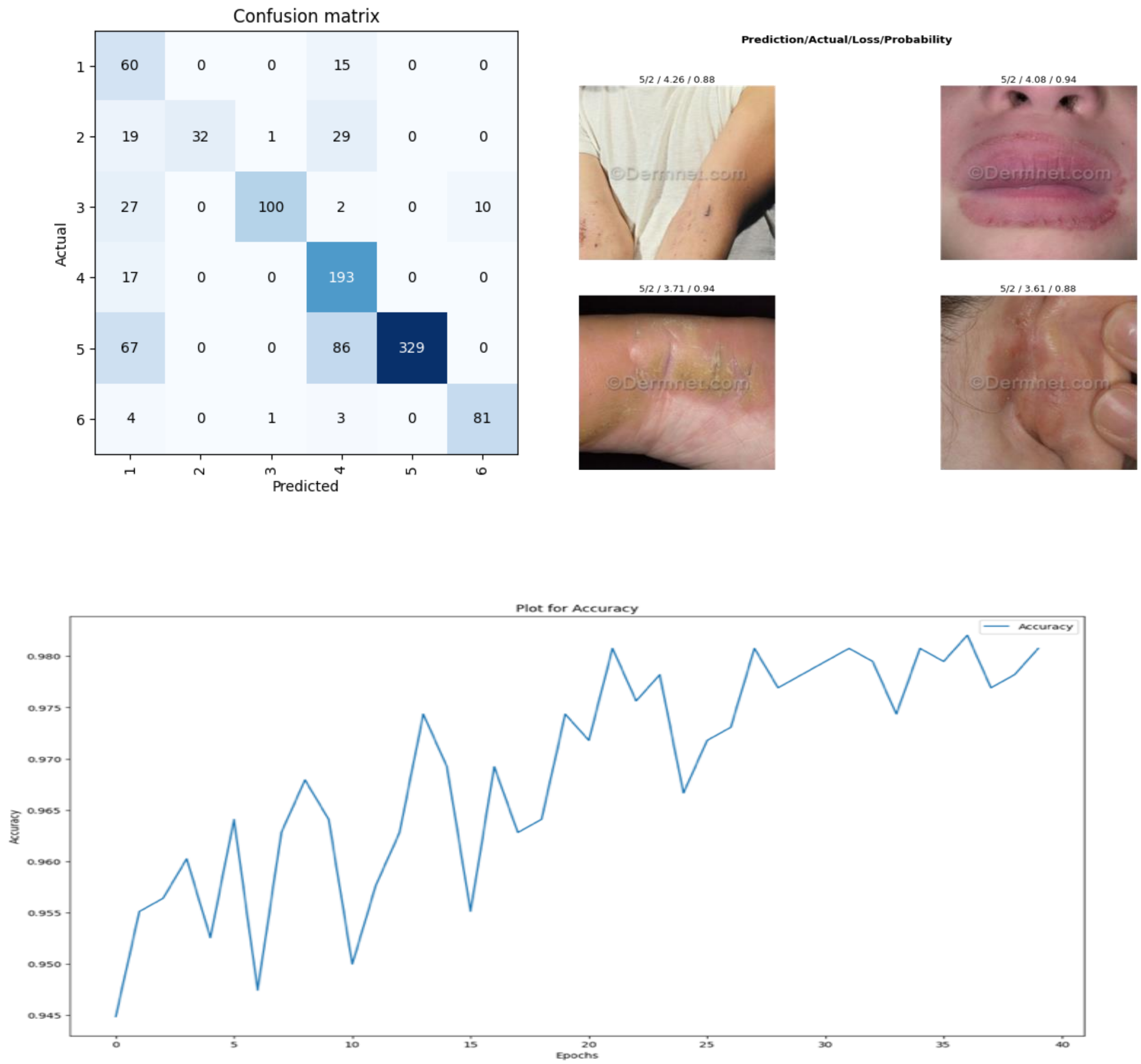

Fig. 18. Model Classification Result

Figure 18 reveals the result of the model's classification of images. The confusion matrix above is a multi-class classification with 6 classes. Each class represents the skin types in the dataset (chickenpox, monkeypox, warts, eczema, corns, normal). The matrix shows the number of instances for each class that was classified correctly and incorrectly by the model. For example, the values in class 1 means there were 75 actual instances of class 1 in the test data, and the model predicted 15 of them as class 4 and 60 of them as class 1.

The Plot for Accuracy shows that the more epochs that you run for the model, the accuracy of the prediction goes higher as well. This is because the model learns from its mistakes and steadily improves with each epoch.

4.2.3 Prediction Results

Table 6. Rate of Skin Disease Prediction

Skin Disease	Accuracy
Corns	77%
Monkeypox	99%
Chickenpox	99%
Warts	90%
Eczema	85%
Normal Skin	99%
Overall Average	92%

Table 6 shows the prediction results of the training data. The application predicted corns with 77% accuracy. Chickenpox was predicted with 99% accuracy. Eczema was predicted with 85% accuracy. Monkeypox was predicted with 99% accuracy. Meanwhile, warts were predicted with 88% accuracy. The overall average of the percentages would be 92%

Table 7. Accuracy Before and After Augmentation

Skin Disease	Before Augmentation	After Augmentation
Corns	75%	77%
Monkeypox	92%	99%
Chickenpox	95%	99%
Warts	87%	90%
Eczema	68%	85%

Table 7 presents the accuracy of the model when it was trained without the application of data augmentation, and when it was trained with the application of data augmentation. Results show that the model improved its accuracy in predicting the skin diseases when data augmentation was applied.

4.3 Result Comparison

Table 8. Comparison of Results

Research Title	Date	Technique	Accuracy
Skin Disease Identification Through Classification and Segmentation Using Deep Learning Techniques	March 2023	ResNet-152	Overall: 92%
A Deep Learning Based Framework for Diagnosing Multiple Skin Diseases in a Clinical Environment	April 2021	EfficientNet-b4 Inception-v3	94%
A machine learning approach for skin disease detection and classification using image segmentation	(n.d)	SVM KNN DT	SVM: 95% KNN: 94% DT: 93%
Skin Disease Detection for Kids at School Using Deep Learning Techniques	July 2022	VGG19	99%
Skin Disease Detection Using Deep Learning	2022	ResNet-50 SVM Softmax classifier	87%

Table 8 shows the comparison between this research study and other related skin disease detection studies. The first study entitled *Skin Disease Identification Through Classification and Segmentation Using Deep Learning Techniqu*e used ResNet-152. The accuracy for Monkeypox prediction was 99%, Chickenpox was 99%, Warts was 81%, Eczema was 89%, and Corns was 75%. The accuracy for classifying normal skin was 99%. Calculating the average accuracy percentage for these six categories yielded a value of 92%. The second study is *A Deep Learning Based Framework for Diagnosing Multiple Skin Diseases in a Clinical Environment*. The proponents of this study used EfficientNet-b4 and Inception-v3 to classify skin diseases. The accuracy of the model was 94%. The third study is entitled *A Machine Learning Approach for Skin Disease Detection and Classification using Image Segmentation*. The technique they used to classify and detect skin disease were SVM, KNN, and DT. SVM can detect and classify skin disease by 95% while KNN can detect by 94% and DT by 93%. The next study is entitled *Skin Disease Detection for Kids at School Using Deep Learning Techniques*. The technique they used is VGG19 and it can detect skin disease by 99%. The last study is entitled *Skin Disease Detection Using Deep Learning* and they used ResNet-50, SVM and Softmax Classifier to detect skin disease. The study resulted to have 87% detection accuracy.

5. Conclusion

Based on the results of the study, the researchers have successfully developed a responsive web-based application with the integration of a pre-trained model made by a convolutional neural network architecture algorithm. The model that the researchers have developed showed promising results in classifying the group of viral skin infections that are in the dataset. It can predict the actual data with 80-90+% accuracy. When it is fed with test images, the model can classify the correct skin infection. However, there are also cases wherein the prediction rate drops down to as low as 60% when images are not of a good quality and other factors such as blurriness and the subject being too far from the camera; The classifier can sometimes be inaccurate due to that reason. The researchers also achieved visualizing the model result through a responsive web-application and have successfully used FastAI deep learning library to develop the predictive model. When the developed predictive model in this study is compared to the models used in previous research, the predictive model can compete with the results due to the accuracy being just as similar to the previous research. The difference of this model compared to previous research is the type of skin diseases that it can classify. Some models of previous research can classify skin cancer images, and some models can only classify one type of skin disease. The predictive model that the researchers have developed is able to classify five types of skin diseases, namely Monkeypox, Chickenpox, Eczema, Corns, and Warts.

This research topic can help in the early identification and early treatment of diseases before they spread because most skin diseases can spread easily with physical contact. This may help in the detection of skin diseases during a public health event and in some rural areas where there is already a lack of basic medical facilities. All in all, by using this technology, it is hoped that skin diseases can be detected and treated early before they can spread and cause further harm.

5.1 Recommendations

The study only focused on classifying five skin diseases particularly monkeypox, chickenpox, warts, eczema, and corns using ResNet-152 as the pre-trained model. However, there are other approaches to classifying skin diseases. These approaches could be used in future studies about classifying skin diseases.

Based on the results of this study, the researchers recommend that the future researchers gather more images that are of high quality and are more focused on the appearance of skin disease for the dataset. This will help lessen the inaccurate prediction of the model. The researchers would also recommend trying to utilize deep learning in detecting other skin diseases that are not mentioned in this study to further contribute to the growth and development of technology in the medical field and dermatology.

References

Ahammed, M., Al Mamun, Md., Uddin, M.S. (2002). A Machine Learning approach for skin disease detection and classification using image segmentation. *Healthcare Analytics,* Volume 2, 2022, 100122, ISSN 2772-4425, https://doi.org/10.1016/j.health.2022.100122.

Alghieth, M. (2022). Skin disease detection for kids at school using deep learning techniques. *International Journal of Online and Biomedical Engineering (IJOE)*, vol. 18, no. 10, 26 July 2022, pp. 114–128, 10.3991/ijoe.v18i10.31879.

Benedetti, J. (2023). Diagnostic tests for skin disorders. MSD Manual Professional Edition. https://www.msdmanuals.com/professional/dermatologic-disorders/approach-to-the-dermatologic-patient/diagnostic-tests-for-skin-disorders

Cleveland Clinic. (n.d.). *Warts: HPV, causes, types, treatments, removal, prevention*. https://my.clevelandclinic.org/health/diseases/15045-warts#management-and-treatment

Codella, N.C.F, Gutman, D., Emre Celebi, M., Helba, B., Marchetti, M. A., Dusza, S.W., Kalloo A., Liopyris, K., Mishra, N., Kittler, H., & Halpern, A. (2018). Skin lesion analysis toward melanoma detection: A challenge at the 2017 International symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC), 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 2018, pp. 168-172, doi: 10.1109/ISBI.2018.8363547.

Darapaneni, N., Sahni, B., Paduri, A. R., Jain, S., Mohamed, S., Banerjee, A., & Chakrabarti, S. (2022). Early detection of skin cancer - Solution for ddentifying and defining skin cancers using AI. *Interdisciplinary Research in Technology and Management (IRTM)*, pp. 1-6, doi: 10.1109/IRTM54583.2022.9791676.

Data analysis and visualization in Python for ecologists: Starting with data. (2023, May 8). https://datacarpentry.org/python-ecology-lesson/02-starting-with-data

De Goma, J. C., & Devaraj, M. (2020b). Recognizing common skin diseases in the Philippines using Image Processing and Machine Learning Classification. https://doi.org/10.1145/3418688.3418700

De Goma, J. C., & Devaraj, M. (2020b). Recognizing common skin diseases in the Philippines using Image Processing and Machine Learning Classification. https://doi.org/10.1145/3418688.3418700

Dev-Site. (2022). *A Guide to the kinds of skin diseases and how to treat them*. Makati Medical Center. https://asm2021.makatimed.net.ph/blogs/a-guide-to-the-kinds-of-skin-diseases-and-how-to-treat-them/

Dildar, M., Akram, S., Irfan, M., Khan, H. U., Ramzan, M., Mahmood, A. R., Alsaiari, S. A., Saeed, A. H. M., Alraddadi, M. O., & Mahnashi, M. H. (2021). Skin cancer detection: A review using deep learning techniques. *International journal of environmental research and public health*, 18(10), 5479. https://doi.org/10.3390/ijerph18105479

Guardian News and Media. (2013, June 8). *Ask a grown-up: How did the first person catch chickenpox?*. https://www.theguardian.com/lifeandstyle/2013/jun/08/ask-grown-up-chickenpox#:~:text=than%20human%20beings.-,The%20first%20chickenpox%20viruses%20probably%20emerged%2070m%20years%20ago%2C%20around,viruses%20have%20evolved%20with%20us.

Johnson, J. (2023, April 28). Common skin diseases and conditions. https://www.medicalnewstoday.com/articles/316622

Kshirsagar, P. R., Manoharan, H., Shitharth, S., Alshareef, A. M., Albishry, N., & Balachandran, P. K. (2022). Deep learning approaches for prognosis of automated skin disease. *Life* (Basel, Switzerland), 12(3), 426. https://doi.org/10.3390/life12030426

Mayo Foundation for Medical Education and Research. (2022, April 30). *Common warts*. https://www.mayoclinic.org/diseases-conditions/common-warts/symptoms-causes/syc-20371125

Mayo Foundation for Medical Education and Research. (2022, December 16). *Corns and calluses*. Retrieved March 30, 2023, from https://www.mayoclinic.org/diseases-conditions/corns-and-calluses/symptoms-causes/syc-20355946

Mendoza, I. F. C. (2023, February 7). 386 cases of hand-foot-and-mouth disease recorded in Davao Region from Jan. 1-21. *SUNSTAR*. https://www.sunstar.com.ph/article/1952878/davao/local-news/386-cases-of-hand-foot-and-mouth-disease-recorded-in-davao-region-from-jan-1-21

National Eczema Association. (2022, December 2). *What is eczema?*. (2022, December 2). from https://nationaleczema.org/eczema/

Nelson, J. (2020). Why to add noise to images for Machine Learning. Roboflow Blog. https://blog.roboflow.com/why-to-add-noise-to-images-for-machine-learning/

Sarker, I. H. (2021). Machine Learning: Algorithms, real-world applications and research directions. SN Computer Science, 2(3). https://doi.org/10.1007/s42979-021-00592-x

Skin diseases. Skin MD - Dermatologist in Ortigas, Pasig. (n.d.). https://www.skinmd.ph/conditions/skin-disease

World Health Organization. (n.d.). *Monkeypox*. from https://www.who.int/news-room/fact-sheets/detail/monkeypox