FIRST DRAFT

by Tatenda chiraga

Submission date: 12-Jun-2023 12:19PM (UTC+0200)
Submission ID: 2114341390
File name: FIRST_DRAFT_revised.docx (3.2M)
Word count: 13389
Character count: 76870

AUTOMATIC GRADING OF SKIN CANCER LESIONS

BY

TATENDA HILLARY CHIRAGA (R1917221Q)

Submitted in partial fulfillment of the requirements for the degree of

BSc Honors Computer Science

Department of Computer Science

In the

Faculty of Science and Technology

At the

Midlands State University

Gweru

June 2023

Supervisor: **Mr. Mupfiga**

ABSTRACT

The aim of this project is to create a mobile application that uses image analysis and machine learning to classify skin lesions as benign or melanoma (skin cancer) depending on their appearance. To improve patient outcomes, melanoma must be diagnosed correctly and early. However, because the common methods like visual inspection and dermascopy are subjective many skin lesions are incorrectly diagnosed. Therefore, by using machine learning algorithms to analyze images of skin lesions and provide a classification score we can address the problem of incorrect diagnosis and unnecessary invasive procedures. The software allows users to log in, take images of suspicious lesions and instantly receive a classification. The user also receives additional information on the lesion classification and necessary contact details.

.

DECLARATION

I, **Tatenda Hillary Chiraga,** do hereby declare that I am the sole author of the dissertation. I authorize **Midlands State University** to publish this dissertation to other institutions or individuals for the purpose of scholarly research.

Signature………………………………………… Date……………………..

APPROVAL

This dissertation entitled **"AUTOMATIC GRADING OF SKIN CANCER LESIONS"** by **Tatenda Hillary Chiraga** meets the regulations governing the award of the degree of **BSc Honours Computer Science of** the **Midlands State University**, and is approved for its contribution to knowledge and literary presentation.

Supervisor's signature…………………………… Date………………….

ACKNOWLEDGEMENTS

I want to thank my supervisor for all of his help and advice in putting my dissertation together. Throughout the duration of this assignment, I learned a great deal. I also want to express my gratitude to my sisters Nyasha as well as Sandra for their unwavering support. My final thank you goes out to all of my fellow students who I bounced ideas off of since their insightful remarks helped me to get better at thinking.

DEDICATION

I dedicate this project to my mom and dad, I hope I make them proud. I would also like to dedicate this project to the almighty.

Table of Contents

LIST OF ACRONYMS

TN…………………………………………………… TRUE NEGATIVE

TP……………………………………………………TRUE POSITIVE

FN…………………………………………………… FALSE NEGATIVE

FP…………………………………………………….. FALSE POSITIVE

CNN…………………………………………………. CONVOLUTION NEURAL NETWORKS

FAQs…………………………………………………FREQUENTLY ASKED QUESTIONS

ROI…………………………………………………… RETURN OF INVESTMENT

IDS……………………………………………...… INTRUSION DETECTION SYSTEM

VPN…………………………………………………VIRTUAL PRIVATE NETWORK.

GPU……………………………………………. ….. GRAPHICIS PROCESSING UNIT

EHR…………………………………………………ELECTRONIC HEALTH RECORDS

DWT………………………………………………… DISCRETE WAVELET TRANSFORM

DFD…………………………………………………DATA FLOW DIAGRAM

LIST OF TABLES

LIST OF FIGURES

Chapter 1: Introduction

1.1 Introduction

Skin lesions are a type of irregularity or injury to the skin that alters its texture, color, and appearance. Many conditions, including infections, allergies, autoimmune disorders, genetic predispositions, and environmental factors including sun exposure, can result in these lesions. Papules, nodules, plaques, pustules, cysts, and ulcers are a few examples of common skin lesions. Particularly for those who are not familiar with dermatology, it can be challenging to recognize and categorize certain skin disorders just on their appearance. The treatment and diagnosis of skin disorders like melanoma depend heavily on the identification and classification of these lesions. Melanoma is a kind of skin cancer that develops from melanocytes, meaning cells that produce pigment. Skin lesion classification is a crucial component of dermatology. This chapter will provide a quick overview of the methods that are currently being used to classify skin lesions. These methods include dermatoscopy, visual examination, and biopsy. We will also quickly describe the drawbacks of the present and prior methods and how the suggested solution will get over such limitations. The main goal of the proposed application is to develop a machine learning-based automated technique for grading lesions. The software will classify the lesions it analyses and determine whether they are cancerous or benign. The automatic skin cancer grading app will collect data when users upload images of their lesions. Pre-processing will then be applied to the image. Pre-processing, also known as input processing, entails altering the image data by scaling the image if necessary and removing undesirable characteristics like hairs, veins, and skin color. The data is then classified according to parameters such as accuracy, sensitivity, and specificity into their deserving categories. Also, the application will offer information on skin cancer and essential dermatological need to know facts. Moreover, access will be made to dermatologists' contact details so that users can follow up with healthcare professional for further assistance.

1.2 Background

Correct categorization is crucial for both treatment and diagnosis. Over the years, several techniques have been employed to categorize skin lesions. The most typical categorization technique is visual examination. It entails a naked eye examination on the appearance, size, shape, color, and texture. These characteristics allow for the classification of the lesion as either malignant or benign. It can only be done by medical professionals like general practitioners and dermatologists and does not need any particular equipment. The process is non-invasive and efficient. Visual examination, however, has several limitations. It is based on the examiner's knowledge and experience and is subjective. In order to accurately classify skin lesions, visual inspection might be performed in conjunction with other techniques. Dermatoscopy is a different technique that may be used to categorize skin lesions. A portable instrument called a dermoscope is used in the non-invasive procedure of dermatoscopy to magnify and highlight skin lesions. Dermatologists may see things beneath the skin's surface that are invisible to the naked eye, such as blood arteries, colour patterns, and other characteristics. These characteristics allow for the classification of the lesion. Dermotoscopy can identify early and subtle change in the lesions and is far more accurate than visual assessment. Moreover, this technique lessens the necessity for aggressive skin biopsies. Dermatoscopy is a technique that may be used to track the development of skin lesions through time. Yet this approach needs special expertise. It takes a lot of time, and not all healthcare facilities always have access to the necessary instruments. When used in conjunction with other techniques, dermatoscopy delivers a more accurate classification. Biopsy is a medical procedure in which a small sample of the skin lesion tissue is removed for a histopathological examination. Skin lesions can be classified according to their histological characteristics. Visual examination and dermatoscopy are thought to be less reliable than biopsies. Yet, a biopsy can be painful, leave scars, or even infect you. Particular tools and knowledge are required for a successful biopsy to be carried out. Furthermore, this method is costly and time-consuming, as seen in cases where it might take days for a patient to receive their results. The classification of skin lesions using machine learning is a promising method. Images of the lesion are analysed using computer algorithms. The method can build a model for the categorization of skin lesions from a large dataset. Machine learning is precise and objective. Moreover, it lessens the need for invasive biopsies.

1.3 Problem definition

- Not everyone has free and easy access to dermatologists
- Due to low turnouts of patients at dermatologists, researchers have limited samples of data for clinical research
- Third world health care infrastructure do not always support data exchange between institutions therefore making outbreak predictions less likely
- Lack of automation is not always efficient and fast which is necessary when dealing with cancer
- Subjectivity: the other methods used are subjective and vary depending on the experience and expertise of the healthcare provider.

1.4 Aim

- Classify skin lesions accurately and efficiently
- Enable users who do not have access to dermatologist to access the platform
- To support data exchange in third world health infrastructure's
- To educate user on useful facts about skin lesions

1.5 Objectives

- Allow users to upload images of different formats and sizes
- Save the uploaded images to a storage system
- Analyze the images using machine learning algorithms and provide categorized accurate results
- Display additional information on the categorized images such as related articles and FAQs
- Provide a contact page for users to send inquiries' or feedback

1.6 Methods and instruments

- Dataset

- Python
- React native
- Yolov5
- Expo

1.7 Delimitation and limitations

- The system does not offer 100% accuracy
- The system does not offer the empathy that goes with human contact
- Encourages hypochondriasis – hypochondriasis is **a condition in which a person is excessively and unduly worried about having a serious illness**
- Bias: Machine learning models can perpetuate bias if the training data is not diverse. This can lead to inaccurate classification for certain groups.
- Limited availability of high quality data: this is critical especially when dealing with rare or unusual skin lesions where limited data can hinder the performance of machine learning models.

1.8 Justification and rationale

Machine learning cannot replace doctors and specialist but has a lot of benefits that go with it.

- **Outbreak prediction**- this can be very important for third world countries whose infrastructure does not support data exchange between health institutions.
- **Data collection** –the data collected from the uses can be used in clinical research which can lead to a cure being produced
- **Melanoma** is a deadly cancer therefore efficient grading is very crucial
- **Efficiency**- the whole process of uploading the image and processing it takes minutes whereas making an appointment and seeing a dermatologist is a lengthy process

1.9 Work Plan

A work plan is a guide that shows the project activities and their timelines. Below is a work plan for the proposed system.

Table 1.1: Work plan

Task	**Duration**	**Start date**	**End date**
Planning	1 week	09-11-2022	16-11-2022
Systems analysis	1 week	17-11-2022	24-11-2022
System design	2 weeks	25-11-2022	09-12-2022
Development	4 weeks	10-12-2022	07-01-2023
Integrating and testing	3 weeks	08-01-2023	29-01-2023
Implementation	2 weeks	30-01-2023	13-02-2023
Operations and research	3 weeks	14-02-2023	07-03-2023

1.9.1 Gantt chart

Week	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Planning																
System analysis																
System design																
Development																
Integration and testing																
Implementation																

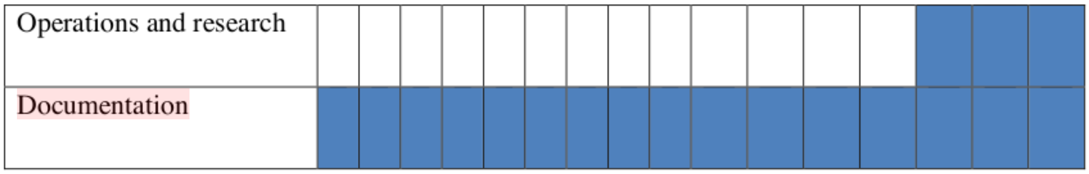


Figure 1.1: Gantt chart

1.10 Conclusion

This chapter summarizes the operation of the automated grading of skin lesion application. We found that the system is efficient and simple to use and access. Users' data can also be used for other purposes such as outbreak forecasting and clinical research. Furthermore, the chapter emphasizes the shortcomings of the proposed method, specifically the absence of various data sets to assure accurate and fair grading. Moreover, we briefly examined the many ways that have been used to classify skin lesions throughout the years, revealing that each method has merits and limitations. As a result, for correct categorization of skin lesions, a combination of methods should be used. The next chapter will focus on evaluating the current work with the previous one depicting the current implementation that overcome the previous problems and limitations in skin lesion grading.

Chapter 2: Literature Review

2.1 Introduction

Skin cancer is a dangerous and sometimes fatal condition that affects millions of individuals throughout the world. Skin cancer detection is critical for successful treatment, but it can be difficult due to the vast number of skin lesions that must be analyzed. In this literature review, we will look at the numerous methodologies that have been utilized for skin lesion classification, noting the benefits and shortcomings of each. The chapter will also discuss how the information gained from the various procedures of dermoscopy, visual examination, and biopsy may be combined with machine learning to provide a more efficient and accurate classification. By understanding the current state of the art in this area, we can identify opportunities for further research and development of more accurate and reliable automated skin lesion classification systems.

2.2 Visual inspection

During a visual inspection, a healthcare practitioner will carefully examine the lesion. They will observe its size, shape, color, and texture. The medical professional may also take notes on the location of the lesion and whether or not it is causing discomfort or pain. They may also inquire about any recent changes in the look of the lesion or any other symptoms the patient is experiencing. It is critical for the healthcare practitioner to comprehend the lesion in order to establish an accurate categorization and select the best approach to treatment for the patient.

Visual examination is an important technique in dermatology since it may aid in the diagnosis of many skin disorders. Visual inspection is very important in the assessment of skin lesions. Dermatologists can tell the difference between different types of skin lesions depending on their appearance. The ability to differentiate is critical since it aids in determining the proper grading and treatment. Melanoma, for example, is a kind of skin cancer that frequently manifests as a black, irregularly shaped mole. Seborrheic keratosis, a common benign skin growth, often manifests as a raised, waxy, wart-like lesion. Dermatologists can give the best treatment options for their patients by studying the visual features of these skin lesions. Visual examination is a very cost-effective approach of classifying skin lesions. It is not only accurate and dependable, but it also requires simply the knowledge of a trained dermatologist. This indicates that even in the absence of expensive equipment and software, facilities may nevertheless provide high-quality skin lesion classification services. Furthermore, visual inspection provides for a more personalized approach to patient care since the dermatologist may consider the patient's medical history as well as general skin health when establishing a diagnosis. Overall, while automated approaches have their advantages, eye inspection is clearly a vital tool in the field of dermatology. The subjective nature of visual examination is one of its drawbacks. Different dermatologists may perceive the look of a lesion differently, resulting in categorization discrepancies. As a result, certain forms of skin lesions may be misdiagnosed or over diagnosed. Another restriction is the absence of quantitative data from visual inspection. Dermatologists may be unable to precisely quantify a lesion's size, depth, or thickness, which can be key considerations in establishing the severity of the lesion. Furthermore, visual inspection may be ineffective in detecting skin lesions that are not apparent to the human eye. Visual inspection alone, for example, may not reveal preclinical abnormalities that have not yet surfaced.

Dermatologists have traditionally depended on visual inspection when assessing skin lesions. However, with the introduction of machine learning, the categorization of skin lesions has become more efficient and accurate. Machine learning algorithms may be trained to analyze thousands of images of skin lesions, discover patterns, and classify them as benign, malignant, or unclear. When combined with eye inspection, machine learning algorithms can help dermatologists make more accurate classifications, especially in difficult instances. However, it's essential to note that machine learning algorithms are not a substitute for human dermatologists.

2.3 Dermoscopy

Dermatologists can use dermoscopy to identify and analyze the many aspects of skin lesions that are not visible to the naked eye. These characteristics include the color, texture, and pattern of the lesion, as well as the presence of hair, scales, or crusts. Dermatologists can distinguish between benign and malignant skin lesions with great accuracy by studying these traits. This is critical because early diagnosis and treatment of malignant skin lesions can improve patient outcomes and lower the risk of complications. A dermatologist will apply a little amount of oil or alcohol to the skin during a dermoscopy examination to eliminate reflections and improve the image. This is done to improve the examination's accuracy and the possibility of discovering any potential skin abnormality. Furthermore, the dermatoscope is put on the skin and rotated about to see the lesion from various angles, providing a more complete assessment of the region. The dermatologist can then notice any anomalies in the skin structure or color and provide a professional diagnosis based on their findings. With this information, the dermatologist can then recommend a suitable course of therapy for the patient, which may involve additional testing or referral to a specialist if necessary. Dermoscopy is becoming an increasingly significant diagnostic technique in dermatology because it allows physicians to more accurately analyze skin lesions and identify probable skin malignancies at an earlier stage, resulting in better patient outcomes and overall health. Dermatologists may also utilize this technology to track the evolution of skin lesions over time, allowing them to make better educated treatment decisions and assess the success of various therapies. Dermoscopy identifies characteristics, patterns, and structures of the skin that are not apparent to the human eye and are crucial in identifying and classifying skin illnesses. One of the main limitations of dermoscopy is that it is operator-dependent. The accuracy of dermoscopy depends on the skill and experience of the operator, and is subject to human error. The operator's level of expertise and training can greatly affect the reliability and accuracy of the diagnosis. In addition, even experienced operators may face challenges when interpreting certain skin lesions, such as those that are atypical in appearance or located in difficult-to-examine areas. Therefore, it is important to have well-trained and experienced operators when performing dermoscopy procedures in order to minimize the risk of inaccurate results.

Furthermore, the operator's equipment and environment can also affect the accuracy of dermoscopy. Poor lighting or a suboptimal imaging device can lead to substandard images, which may affect the operator's ability to make a correct classification. Therefore, it is important to ensure that the operator uses high-quality equipment and works in an appropriate environment to optimize the accuracy of dermoscopy. Lastly, it is worth noting that the operator's interpretation of the dermoscopic image is also influenced by their knowledge of the lesion and its clinical context. Dermoscopy is not a diagnostic tool in itself, but rather one of several tools available to aid in the diagnosis of skin lesions. Therefore, the operator's interpretation of the dermoscopic image must be used in conjunction with a thorough clinical examination and relevant patient history to arrive at an accurate diagnosis. While dermoscopy is a useful tool in the diagnosis of melanoma, it has its limitations. Dermoscopy can only provide information about the surface features of a lesion and cannot provide information about the depth of the lesion. Therefore, dermoscopy alone may not be sufficient to diagnose melanoma accurately. It is important to note that dermoscopy is just one of many tools that can be used to diagnose melanoma. Other tools, such as biopsy and histopathology, can provide additional information about the lesion and help to confirm a diagnosis of melanoma. Additionally, new technologies and techniques are constantly being developed to improve the accuracy of melanoma diagnosis. For example, confocal microscopy is a non-invasive imaging technique that can provide high-resolution images of the skin and can be used to diagnose melanoma with greater accuracy than dermoscopy alone. Also, it is important to consider the context in which dermoscopy is being used. Dermoscopy may be less effective in diagnosing melanoma in certain populations, such as those with darker skin tones, due to differences in the appearance of lesions. In these cases, additional diagnostic tools may be necessary to accurately diagnose melanoma.

Overall, while dermoscopy is a valuable tool in the diagnosis of melanoma, it is important to recognize its limitations and to use it in conjunction with other diagnostic tools to achieve the most accurate diagnosis possible. Dermoscopy can utilize artificial intelligence and machine learning algorithms to analyze and classify skin lesions based on dermoscopy images. This technique overcomes the traditional method's limitations by providing a more accurate and efficient diagnosis of skin diseases.

2.4 Biopsy

Mayo Clinic (2019) defines a biopsy as a simple procedure in which a sample of the skin is removed and tested in a laboratory. There are different types of biopsies which include punch biopsy, shave biopsy and excisional biopsy. The choice of biopsy depends on the size and location of the lesion.

A circular instrument is used to remove a tiny cylinder of tissue from the lesion during a punch biopsy. Tissue from skin layers such as the epidermis, dermis, and the top layer of fat under the skin may be included in the sample. Following the treatment, the patient may require stitches to close the incision. This form of biopsy is frequently utilized for tiny lesions or lesions on the face, neck, or scalp. According to a Mayo Clinic article, a shave biopsy is a process that includes scraping the surface of the skin using an instrument similar to a razor. The process collects a tissue sample from the skin's top layer. This type of biopsy is often used for superficial lesions and little to no stitches are requires after the process. Mayo Clinic (2019) defines excisional biopsy as the process of completely removing a solitary skin lesion or area of irregular skin. The excised skin sample may contain a border of healthy skin as well as the skin's deeper layers. Following this operation, stitches may be required. The tissues samples are delivered to the laboratory for analysis after being collected using various ways.

Dermatologists frequently use visual examination and dermoscopy to diagnose and classify skin lesions. These approaches, while valuable, are not always correct. This is due to the fact that there are several varieties of skin lesions, some of which might be difficult to differentiate from others based just on appearance. A biopsy, on the other hand, can give a more precise procedure. Doctors can determine the kind of lesion and offer critical information about its severity and treatment choices by obtaining a tiny sample of the lesion and analyzing it under a microscope. While visual inspection and dermoscopy are useful methods for assessing skin lesions, biopsies are frequently required to establish a final classification and assure proper treatment. One limitation of biopsy is that it may not sample the entire lesion, which can be large and irregularly shaped. As a result, a biopsy may only sample a tiny fraction of the lesion, resulting in an incomplete or inaccurate prognosis. However, biopsies remain an important technique in dermatology, and advances in technology have allowed for more accurate sampling through methods such as Mohs micrographic surgery, which involves the removal of thin layers of skin in layers until the entire lesion is removed. Furthermore, clinicians may use imaging methods such as ultrasound to offer a more comprehensive view of the lesion and to guide the biopsy operation for a more precise classification. Another drawback of biopsies is that they can be intrusive and uncomfortable, causing discomfort and possibly scarring in people who have the operation performed. It is important to note that this discomfort may prevent some patients from seeking medical attention or following through with recommended treatments. Advances in medical technology and techniques have helped to minimize the discomfort and scarring associated with biopsies, and healthcare providers can work with patients to provide support and guidance throughout the procedure and recovery process. In addition, biopsies can be expensive, especially if multiple lesions need to be analyzed. This cost can be a barrier for some patients who may not have insurance or the financial means to cover the procedure.

Biopsies are an important classification tool in skin grading because they allow clinicians to acquire tissue samples for additional study. Biopsies may now be analyzed more efficiently and accurately thanks to the advancement of machine learning, which can assist in the proper classification and treatment of many skin disorders. On huge datasets of biopsy findings, machine learning algorithms may be trained to discover patterns and make predictions regarding the class of skin lesions. A machine learning model, for example, may be created to reliably classify lesions based on biopsy data. Medical practitioners may be able to diagnose cancer lesions sooner and establish more effective treatment strategies by using machine learning. Additionally, machine learning can aid in the processing of complex biopsy data. Biopsies may give a great deal of information beyond determining if a lesion is benign or cancerous, and can even analyze genetic abnormalities in cells. Machine learning methods can be used to detect these mutations and anticipate how the lesions will grow. Overall, the utilization of biopsies in conjunction with machine learning has the potential to transform the medical industry. Machine learning is projected to become an increasingly more potent tool in the classification, treatment, and prevention of skin diseases as technology advances.

2.5 Limitations of the current systems

In order to successfully apply machine learning algorithms for skin lesion classification, it is crucial to have a comprehensive and diverse dataset of skin lesion images. This dataset should be carefully curated and contain a wide range of skin diseases, ensuring that the machine learning algorithm is able to accurately identify and classify all potential skin conditions. It is recommended that the dataset be large enough to include a sufficient number of images for each disease, in order to provide the machine learning algorithm with enough examples to learn from. Once the dataset has been compiled, it is then divided into separate training and testing sets. The training set is used to train the machine learning algorithm, allowing it to learn from the images and develop the ability to accurately classify skin lesions. The testing set is then used to evaluate the performance of the machine learning algorithm, ensuring that it is able to accurately identify and classify skin lesions that it has not seen before. By carefully selecting and curating a comprehensive dataset of skin lesion images and using machine learning algorithms, it is possible to improve the accuracy and efficiency of skin lesion classification, ultimately leading to better diagnosis and treatment of various skin diseases.

- Operator dependence: Interpreting dermoscopic characteristics takes training and skill, and the accuracy of dermoscopic classification is strongly dependent on the ability and experience of the healthcare professional doing the examination. Inexperienced or untrained users may misunderstand dermoscopic characteristics or make classification mistakes, resulting in incorrect diagnosis.
- Dermoscopy instruments and settings might differ, resulting in differences in the appearance of dermoscopic characteristics. Different devices, lighting settings, and magnification levels can impact lesion visualization, and there is no standardized dermoscopy system or settings, which can affect dermoscopic categorization accuracy and consistency. These gaps can be filled by machine learning's capacity to learn from experience. Machine learning algorithms improve their classification of skin lesions as they analyze more data, resulting in higher accuracy over time.
- Visual inspection and dermoscopy may have limitations in certain populations, such as patients with darker skin types or specific dermatologic conditions. Dermoscopic characteristics may be less obvious or varied in various groups, making precise classification difficult. Machine learning algorithms, on the other hand, may be taught on labelled picture datasets available from a variety of sources, including public databases and medical institutes.
- Dermoscopy equipment, doctor consultations, and biopsies can all be expensive. Access to dermoscopy equipment and qualified healthcare personnel may be limited in certain healthcare settings, particularly in resource-constrained locations like the rural areas, affecting the availability and usage of dermoscopy for skin lesion classification. Machine learning algorithms may be used to categorize skin lesions, reducing the requirement for

professional dermatologists and making the procedure more cost-effective and readily available.

- Invasiveness: A biopsy is an invasive technique in which a part of the skin lesion is removed for evaluation. This can result in discomfort, suffering, and potential complications such as bleeding, infection, or scarring. Patients may be apprehensive to undergo biopsy in some situations owing to the intrusive nature of the operation. Machine learning algorithms might possibly determine which skin lesions are benign and which are cancerous without the need for a biopsy by properly identifying skin lesions.
- Variability in interpretation: Histopathological analysis of biopsy samples requires experience in identifying microscopic characteristics, and pathologists' assessment of the samples might vary. Inter-observer variability, in which different pathologists read the same sample differently, can have an influence on the accuracy and consistency of skin lesion classification. Large datasets of biopsy photos accurately categorized by dermatologists can be used to train machine learning systems. The machine learning system can then make predictions on new imagery that it has never seen before by learning patterns in existing images. This can assist to lessen variation in interpretation among dermatologists since the machine learning system will identify the picture based on the patterns it has learnt rather than on subjective interpretation.
- Delayed results and false negatives are some of the drawbacks when using biopsies to classify skin lesions. Biopsy results may take several days to weeks to become available. This might cause delays in receiving a definitive diagnosis and commencing appropriate therapy, which can be a problem in urgent or time-sensitive instances. One of the most notable benefits of utilizing machine learning to classify skin lesions is its capacity to analyze large volumes of data fast. Machine learning algorithms can uncover patterns that are difficult to perceive with the human eye by analyzing massive datasets of skin lesion photos, resulting in more accurate diagnoses.
- The issue of security in healthcare is a significant concern that needs to be addressed. Healthcare data is sensitive and confidential, and it is necessary to ensure that it is protected from unauthorized access, manipulation, and theft. To ensure privacy the images are pre-processed and analyzed locally and are never uploaded to an external server

2.6 Conclusion

In conclusion, the literature review suggests that machine learning algorithms can be effective in the classification of skin lesions. Machine learning methods for classifying skin lesions overcome the challenges associated with earlier classification systems in terms of accuracy, security, efficiency, and availability. We may also infer that, while machine learning can help dermatologists make more accurate classifications, it cannot replace a human dermatologist's skill and experience. As a result, it is vital to employ machine learning as a tool to assist in decision-making rather than as a replacement for human knowledge.

Chapter 3: Analysis phase

3.1 Introduction

The field of healthcare has seen significant advancements with the integration of machine learning techniques. A thorough analysis phase is essential for developing a robust and successful skin lesion categorization application. The analysis phase is important to the project's success since it establishes the groundwork for the development of the machine learning system. By conducting a feasibility study and identifying functional and non-functional requirements, the project can proceed with a clear understanding of the goals and limitations of the project.

3.2 Rationale of the proposed system

- **Accuracy**- Dermatologists are critical in the diagnosis and treatment of skin cancer, and visual inspection is the primary method used for identifying malignant skin lesions. Even the most skilled dermatologists often struggle to differentiate between benign and malignant lesions, resulting in delayed diagnosis and treatment. By finding patterns in skin lesion images that are not visible to the human eye, machine learning algorithms may greatly improve the accuracy and efficiency of skin cancer detection. Machine learning algorithms can uncover critical traits that differentiate between malignant and benign skin lesions by analysing massive datasets of skin lesion images, allowing doctors to make more accurate diagnosis. These algorithms may also help identify skin lesions that a human dermatologist would miss, which is especially valuable in locations with limited access to dermatologists or long wait periods for appointments. Furthermore, machine learning algorithms may be programmed to learn and increase their accuracy over time, making them a useful tool for dermatologists in grading skin cancer. This can lead to earlier detection, more effective treatment, and improved patient outcomes in the long run.
- **Efficiency**- Machine learning in skin lesion grading has the potential to transform skin cancer classification and treatment. We can guarantee that patients obtain fast and correct diagnoses by employing algorithms to aid dermatologists in their diagnosis, which can ultimately save lives. The entire process of uploading and analysing the image takes minutes, whereas setting an appointment and seeing a dermatologist takes time.

- **Outbreak Prediction-** Machine learning-based classification of skin lesions can help predict and avoid infectious disease epidemics. We can analyze large volumes of data from skin lesion photos using machine learning algorithms to discover patterns that signal the presence of a certain illness. This is especially significant in third-world nations whose infrastructure does not allow for data interchange across health-care facilities. Using this technology, we may detect epidemics early and intervene rapidly to prevent disease transmission. Also, machine learning algorithms can aid in the detection of new or developing illnesses**.** We can uncover new diseases and swiftly build a reaction to prevent the outbreak from spreading by analyzing skin lesion images and discovering patterns that are not currently related with any recognized disease.

- **Non Subjective**- Conventional methods of classifying skin lesions rely on dermatologists' expertise, who visually assess the lesion and diagnose it based on their experience and knowledge. However, this procedure is susceptible to human error and can result in misdiagnosis. Machine learning algorithms, on the other hand, are trained on enormous datasets of skin lesion images, allowing them to detect minute variations in lesions that the naked eye may miss. This leads to a more accurate classification and decreases the possibility of misdiagnosis.
- **Data collection** – The data gathered from users can be used in clinical research, perhaps leading to the development of a cure. Dermatologists may also rapidly and efficiently gather vast amounts of data, which can subsequently be utilized to build more accurate diagnostic tools and treatment regimens.
- **Improved Accessibility**-Access to dermatological expertise may be limited, particularly in rural or underprivileged locations with a scarcity of dermatologists. Machine learning technologies can fill this need by making skin lesion analysis solutions that are both accessible and inexpensive. This technology allows users to obtain early assessments and suggestions remotely, possibly saving both patients and healthcare professionals time and resources.
- **Early Detection and Intervention**- Skin cancer, such as melanoma, can be fatal if not detected and treated early. By recognizing certain patterns or traits suggestive of cancerous development, machine learning algorithms can help in the early diagnosis of possibly malignant skin lesions.

Early intervention may greatly improve patient outcomes and increase the likelihood of treatment success.

- **Reduction of Diagnostic Errors**-Human bias, weariness, or a lack of experience can all contribute to diagnostic mistakes. Machine learning models can help dermatologists make decisions by providing an extra layer of assistance. The possibility of mistakes and oversights may be reduced by incorporating machine learning into the diagnostic process, leading to increased accuracy and patient care.
- **Efficient Triage and Prioritization**-Dermatologists frequently deal with a high volume of patients and skin lesion situations, necessitating effective triage and priority. Machine learning systems can assist dermatologists grade and prioritize cases depending on their level of concern, allowing them to better focus their time and resources. Cases that are urgent or suspicious can be marked, ensuring that high-risk patients receive immediate treatment.

3.3 Feasibility analysis

A feasibility study is essentially an assessment of the practicality of a given project plan or approach. This is accomplished by an examination of technological, economic, social, and operational factors. Such a study seeks to demonstrate if the project can be built within a certain set of targets and time constraints. If the project satisfies the established limits, it is regarded acceptable since it outweighs the shortcomings predicted from its failure.

3.3.1 Technical feasibility

Technical feasibility is defined as the formal process of determining whether a product or service is technically feasible to develop. By analyzing the process, tools, software, hardware, labor, and logistics, technical feasibility helps establish the efficacy of the proposed method.

- **Availability of Data**-The availability and quality of the data used to train the model is critical to the success of any machine learning technique. To train the algorithm for skin lesion categorization, a large dataset of high-quality images is necessary. While various publicly available datasets are available, they may not be a representative of all skin types and conditions. As a result, data collection and curation are important to ensuring the model's robustness and accuracy. The HAM10000 dataset was used in this project. Collaborations with dermatologists or medical institutes may be required to collect consented images of skin lesions when developing an application for a specialized domain. The availability of resources, ethical issues, and legal compliance, such as patient privacy and data protection rules, all influence the viability of data collecting. After obtaining the data, the images must be tagged or labelled with the relevant lesion types or conditions. This might be a time-consuming operation requiring the use of skilled dermatologists. The availability of competent personnel and resources determines the possibility of precise and consistent labelling. Annotation tools such as Roboflow can be utilized. Data augmentation techniques can be used to expand the amount of the dataset artificially and improve the model's generalization. To produce variations, techniques such as rotation, flipping, scaling, and adding noise can be used to existing images. Data quality is critical for developing an efficient skin lesion classification model. Images ought to have good resolution, correctly labelled, and free of artefacts or distortions such as hair. Furthermore, any biases in the dataset, such as imbalances in lesion types or demographic biases, must be considered. It is critical for reliable and fair classification results to ensure data quality and minimize bias. As a result, having a dataset that reflects various skin tones is critical.

- **Complexity of the Algorithm/ reliable machine learning algorithm-**The technological feasibility of skin lesions might be influenced by the complexity of the machine learning method employed to categorize them. Convolutional neural networks (CNNs), for example, have shown considerable promise in image classification applications such as skin lesion grading. Model architecture is crucial for getting high classification accuracy. The feasibility of the project is dependent on selecting an appropriate CNN architecture and tuning its key parameters to obtain the needed accuracy and efficiency. MobileNet is a CNN architecture developed for mobile and embedded devices. MobileNet is purpose-built to be light and efficient, allowing it to function on devices with minimal processing capabilities, such as smartphones. The main idea behind MobileNet is to reduce the number of parameters and operations in the network while still maintaining good accuracy on image classification tasks. It achieves this by using depth-wise separable convolutions, which split the standard convolution operation into separate depth-wise and point-wise convolutions. The depth-wise convolution applies a single convolutional filter to each input channel separately, reducing the number of input channels. The point-wise convolution then applies a 1x1 convolution to mix the reduced channels. By using this combination of depth-wise and point-wise convolutions, MobileNet significantly reduces the computational cost while retaining the ability to learn complex features

- **Model training and Validation**-Training a machine learning model involves optimizing its parameters using the labelled dataset. The viability of training is dependent on the availability of computational resources, such as GPUs, to analyse massive amounts of data and train complicated models efficiently. Additionally, proper validation techniques such as cross-validation or hold-out validation must be used to ensure model generalization and avoid overfitting.
- **Deployment Considerations**-Deploying a machine learning system for skin lesion categorization necessitates careful consideration of a number of technical issues. These include the model's computing needs, interaction with the application or system, scalability to handle real-time demands, and the capacity to manage massive amounts of incoming data. The availability of infrastructure, such as servers or cloud platforms, to fulfil the algorithm's requirements determines the viability of implementation.
- **Interface Design**-Consider the requirements and expectations of the application's target users while designing a user-friendly interface. The interface should be user-friendly, visually appealing, and give clear instructions. It should make it simple for users to input images of skin lesions, display categorization findings, and present them in a way that is comprehensible.

- **Performance Optimization**-Machine learning for skin lesion categorization can be computationally intensive, especially when working with large images or complex models. To create a seamless user experience, the model's performance and the interface's responsiveness must be optimized. Model compression, hardware acceleration, and cloud-based inference can all be used to improve the application's performance.

3.3.2 Economic feasibility

Economic feasibility includes determining the financial risks, costs, and benefits of the new system. Economic feasibility, assesses if time and money are available to create the system. This feasibility study is necessary to determine the cost to be spent as well as the benefit or return expected from the system's use. All of this will be accomplished through the use of cost-benefit analysis tools.

3.3.2.1 Development Costs

These are the costs of developing the application. The cost of recruiting machine learning specialists, software engineers, and designers is included. You may also need to spend money on hardware infrastructure, software licenses, and data collecting. It might take time and money to collect and annotate a big dataset of skin lesion images. You may need to work with dermatologists or pay license fees to gain access to current datasets. The cost will be determined by the amount and quality of the training and validation datasets.

The following table show the development costs

Table 3.1: Development Costs

Description	Cost us$
• Dermatologists fee • Hardware • Software(Roboflow) • Data acquisition fee	600 - 50 600
Total	1250

3.3.2.2 Cost benefit analysis

Table 3.2: Cost Benefit Analysis

	$	$
Benefits		
Consultation fee		20
Reduced labour cost		600
Reduced healthcare cost		2000
Patient throughput		1000
Total		3600
Development cost		
Dermatologists fee	600	
Software	50	
Data acquisition	600	
Operational cost		
Maintenance and repair	600	
Total	1850	
Net costs		**1750**

3.3.2.3 Return on investment

Return on Investment (ROI) Analysis: Estimates the potential revenue and compare it against the development and ongoing costs. Conduct a thorough financial analysis to determine the payback period and expected ROI. This will help assess the economic viability of the application.

ROI $= \frac{(\text{Total Benefits} - \text{Total Costs}) * 100\%}{\text{Total cost}}$

$$\frac{(3600 - 1850) * 100\%}{1850}$$

$$\frac{1750}{1850} * 100\%$$

94.59%

3.3.3 Social feasibility

Social feasibility is the assessment of whether a proposed project or system is acceptable within the social context. Several elements and concerns influence the social feasibility of an application that employs machine learning to classify skin lesions. Here are some factors to consider while evaluating the social feasibility of such an application.

- **Medical Accuracy and Safety**-The foremost concern for any medical application is its accuracy and safety. To achieve reliable and accurate categorization of skin lesions, the machine learning model should be well-trained and validated on various and representative datasets. To test the application's performance against recognized medical standards, rigorous clinical studies and reviews should be carried out.
- **Privacy and Data Security**- The software should emphasize user privacy and safeguard sensitive medical data. Compliance with the data protection act set by POTRAZ, the current telecommunications authority, is critical. Implementing strong encryption, secure data storage, and informed consent processes can help build user trust and address privacy concerns.
- **Ethical Considerations**: It is important to think about the ethical implications of developing and implementing a machine learning application for healthcare reasons. Transparency in algorithmic decision-making, fairness and bias prevention, and careful

management of sensitive information are all part of this. Regular monitoring and audits should be carried out to guarantee that ethical procedures are followed. Required disclosures should be clearly disclosed. The disclaimer that the categorization is not a medical diagnosis should be apparently displayed.

- **User Acceptance and Trust**: The application's social feasibility is heavily influenced by the public's perception and acceptance. Educating users on the system's limitations and capabilities might help them manage their expectations. Engaging with healthcare experts, patient advocacy organizations, and the general public to gather feedback and resolve problems may boost trust and acceptability.
- **Accessibility and Equity**- To guarantee social viability, the application needs to be accessible to a diverse user base, taking into account aspects such as language challenges, literacy levels, and technical accessibility. It is critical to avoid increasing existing healthcare inequities and to work for fair access to the application across demographics and socioeconomic backgrounds.
- **Collaboration with Healthcare Professionals**-Collaboration with dermatologists and other medical experts is essential for properly validating and integrating the application into clinical processes. Involving healthcare practitioners in the development process, performing collaborative research, and creating clear standards for the application's use can all assist to gain credibility and acceptability within the medical community.
- **Regulatory Compliance**- Compliance with appropriate regulatory frameworks and securing requisite permissions is essential for a medical application's social feasibility. Understanding and abiding to medical software rules, such as those imposed by the Postal and Telecommunications Regulatory Authority of Zimbabwe (POTRAZ) is also important.

3.3.4 Operational feasibility

According to Somerville (2016), operational feasibility refers to the ability of a proposed system to be used efficiently and effectively after it has been developed. This is done to determine whether the system is compatible with the everyday operations. In the instance of a machine learning application used to categorize skin lesions, operational feasibility issues are critical to ensuring the project's feasibility and success. Here are some crucial factors to consider when determining operational feasibility.

- **Data availability**- Assess if a sufficient number of high-quality skin lesion data is available and accessible for training and testing the machine learning model. Ample data is required for training accurate and dependable models. Consider the availability of annotated data, as well as the possible difficulties associated with getting and curating such data.
- **Computational resources**- Examine the machine learning model's computing needs. Skin lesion classification models often demand a substantial amount of computer power and memory, especially if deep learning approaches or huge datasets are used.
- **Integration with existing systems**- Consider the skin lesion categorization application's compatibility and interaction with current systems and workflows. If the program is designed for usage in a healthcare setting, it should interact easily with electronic health record (EHR) systems or other relevant medical software. Smooth integration may reduce interruptions and increase user acceptance.
- **User training and support**- Determine the amount of skill necessary to operate and comprehend the skin lesion categorization application's results. Consider the training requirements for users, such as dermatologists or healthcare professionals, to use the program efficiently. Furthermore, consider the availability of technical assistance or documentation to aid users in diagnosing problems or understanding model results.
- **Regulatory and legal considerations**- Understand the regulatory environment and legal requirements for installing a skin lesion categorization application. Medical software may be required to conform with special rules, such as data privacy (e.g., POTRAZ Privacy act) or medical device regulations (e.g., FDA in the United States), depending on the jurisdiction. To avoid legal difficulties, ensure that the application corresponds to the appropriate standards and rules.
- **Maintenance and updates**- Take into account the application's continual maintenance and updates. Machine learning models may need to be retrained on a regular basis as new data becomes available or if the model's performance deteriorates over time. Evaluate the app's viability, including the availability of resources for monitoring, updating, and improving the model as needed.
- **Scalability-** Examine the application's capacity to scale effectively. As the number of users or the amount of data rises, the application should be able to handle the increased demand

without experiencing severe performance issues. To accommodate possible expansion, assess the scalability of both the computing infrastructure and the machine learning methods.

3.4 Functional requirements

It is described as observable capabilities and functions of a system with the objective of filling the gaps of the existing and previous system

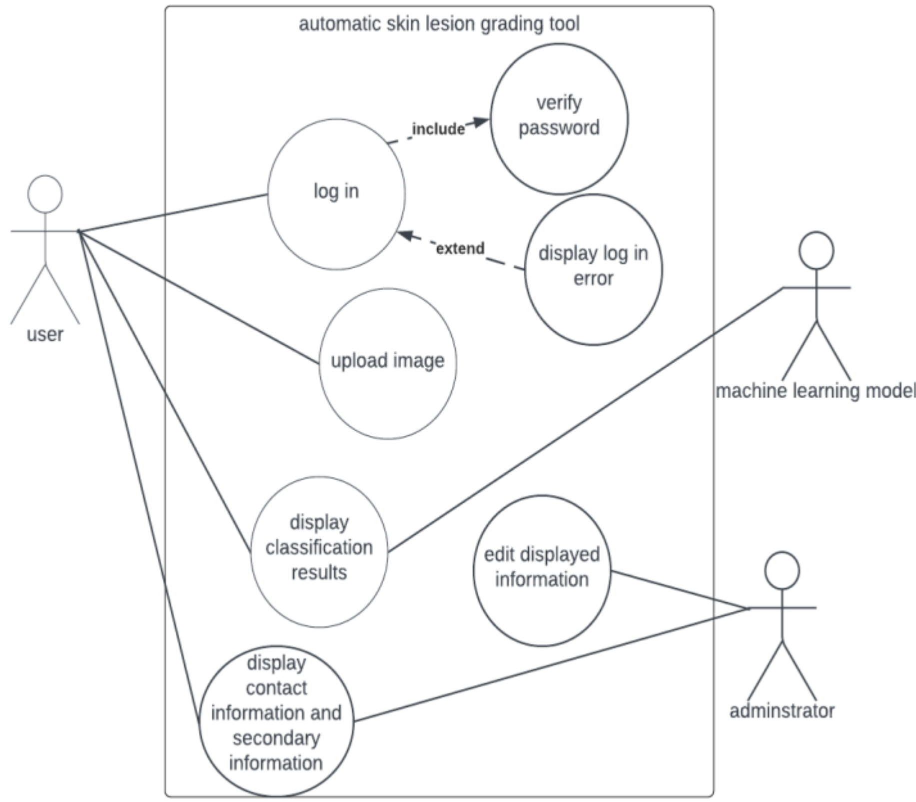


Figure 3.1: Use Case Diagram

- The system must allow user to log in, verify log in details and display login error if the user enter wrong credentials
- The user should be able to upload an image of the lesion
- The machine learning model should generate graded results
- The system should display additional information about the skin lesion classes and dermatologists contact information.
- The administrator should be able to edit additional information and contact details

3.4.2 Non-functional requirements

Skin lesion classification application using machine learning is an essential tool for dermatologists to diagnose skin cancer. This application's success is dependent on its ability to produce accurate and dependable results. This application's non-functional needs are as follows:

- **Speed**- To guarantee that users receive prompt treatment the application must categorize skin lesions within an acceptable time limit. The application should classify a lesion in less than 10 seconds.
- **Scalability-To manage a big amount of images, the application has to be scalable. It should be capable of processing at least 100,000 images each day.**
- **Confidentiality**- The application must protect patient data confidentiality and should not disclose any of the patient's data with individuals who are not authorized.
- **Integrity**- The program must protect the data's integrity by preventing unauthorized changes or tampering.
- **Authentication and Authorization**- **To guarantee that only authorized individuals may access the data, the application must have authentication and authorization protocols.**
- **Ease of Use**- The application should have a user-friendly and easy to navigate interface. The interface should be intuitive and straightforward.
- **Multilingual Support**- The application should support multiple languages to cater to a diverse patient population.
- **Accuracy**- The application must provide accurate results with a high degree of confidence. The application should have an accuracy rate of at least 95%.
- **Availability**- The application must be available 24/7 to ensure that patients can access it at any time.
- **Fault Tolerance**- The application should have a fault-tolerant design that can recover from failures without affecting the availability of the system.
- **Ease of Maintenance**- The application should be easy to maintain and update.
- **Modularity**- The application should be designed in a modular fashion, allowing for easy updates and changes to individual components.

- **Documentation**- The application should have comprehensive documentation that outlines how to maintain and update the system.

3.5 Conclusion

In conclusion, the analysis step of designing a skin lesion classification application using machine learning algorithms is critical to the project's success and effectiveness. Understanding the project's rationale allows us to gain insight into the motive for designing the application. The author also finds that by using an accurate and efficient skin lesion categorization program, healthcare practitioners in dermatology may improve their diagnostic capabilities and provide better patient care. The next chapter will go over the project's design process, with an emphasis on the application interface and how the software and hardware will interact.

Chapter 4: Design Phase

4.1 Introduction

Determining and analyzing the design of the system's modules, elements, and system data will be the focus of this chapter. Ideas are converted into tangible plans and specifications during this stage, laying the groundwork for the implementation and development phases that follows. In this chapter, a data flow diagram and a context diagram will be utilized to clearly describe how the system works. The architectural design, database design, and user interface are all included in this phase and are essential to the development of the complete system.

4.2 Architectural design

Architectural design is the act of identifying a group of hardware and software components, as well as their interactions, in order to establish the basis for a system's development.

The following elements are part of this architecture design in this instance:

- **User Interface**: Users can interact with the application through this component's graphical user interface. Users can log in, upload pictures of skin lesions, check the classification's outcomes, and leave feedback to assist the model's accuracy improve over time.
- **Application Server**: This part is in charge of managing image processing and responding to user requests. The user's uploaded images are received by it, and it then sends them to the machine learning model for processing.
- **Machine Learning Model**: This component uses a trained machine learning model to classify the skin lesions. It receives a skin lesion image as input and outputs a categorization outcome.
- **Database**: This part houses user information, image data, and related classification outcomes. Additionally, it keeps user comments to help the model's accuracy develop over time.

- **Cloud Storage**: This element is utilized to store the dataset of skin lesion images as well as the machine learning model that has already been trained. For the application, it offers great availability and scalability. A good example is Google Colab.

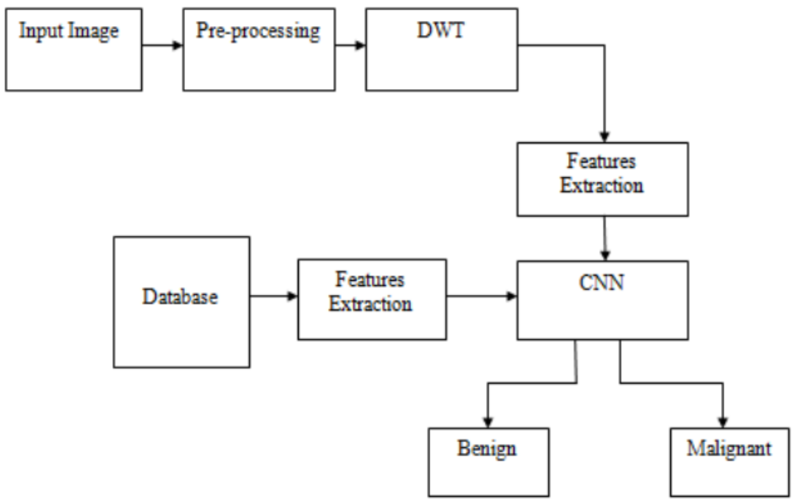


Fig 4.1: Architecture design

4.3 System design

System design is a process that invloves establishing a system's architecture, components, data, interfaces, and modules to satisfy preset requirements. Creating a well-designed system with features like efficacy, dependability, and user-friendliness is the aim of system design.

- User-friendly interfaces that are simple to use and navigate
- Dependability is all about a service's availability when it is required and its reliability in delivering the desired outcomes as the system ought to
- Efficient systems use less time and resources while requiring the least amount of user effort.

How the systems works

- **Data collection and preparation**: Gathering a sizable dataset of skin lesion images is the initial stage. The images must be annotated with the appropriate classification, such as benign nevus, basal cell cancer, or melanoma. To get rid of any irrelevant or duplicate information, the data needs to be pre-processed and cleaned.
- **Feature extraction:** To provide the machine learning algorithm with information, features are first extracted from the images. In order to do this, relevant details from the images must be recognized and selected, such as color, texture, and shape.
- **Model selection and training:** Depending on the issue at hand, a machine learning model is chosen. In order to classify images, convolutional neural networks (CNNs) are frequently employed. The chosen model is then trained using an appropriate training procedure, such as stochastic gradient descent, on the pre-processed data.
- **Model evaluation and validation**: The performance of the trained model is assessed using a different test dataset. The performance of the model is evaluated using metrics like recall, accuracy, and precision.
- **Deployment:** The model can be included in the application when it has been validated and deployed. A user should be able to log in to the application, upload images of skin lesions, have those images processed using the trained model, and then receive a categorization or probability score for each lesion.
- **Administrator responsibilities:** The administrator should have the ability to modify and update information on skin lesions and contact information.

4.3.1 Context diagram

An overview of a process or system that depicts the relationships between the system's components and external entities is known as a context diagram.

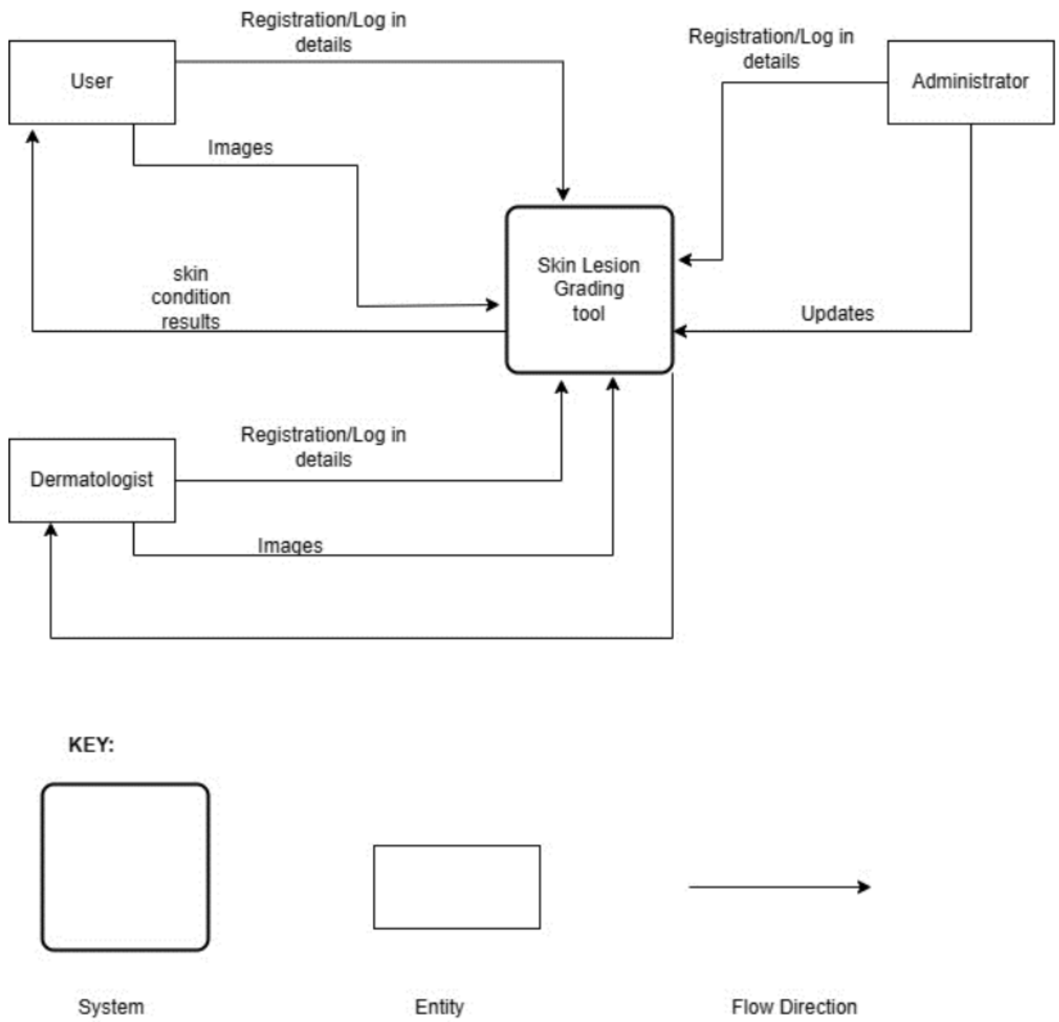


Fig 4.2: Context Diagram

4.3.2 Data flow diagram

A data flow diagram (DFD) is a graphical representation of the flow of data through an information system. It demonstrates the systematic input, processing, and output of information. It can also be referred to as a diagram that depicts the data flow from processes, entities, and data stores from one system component to another.

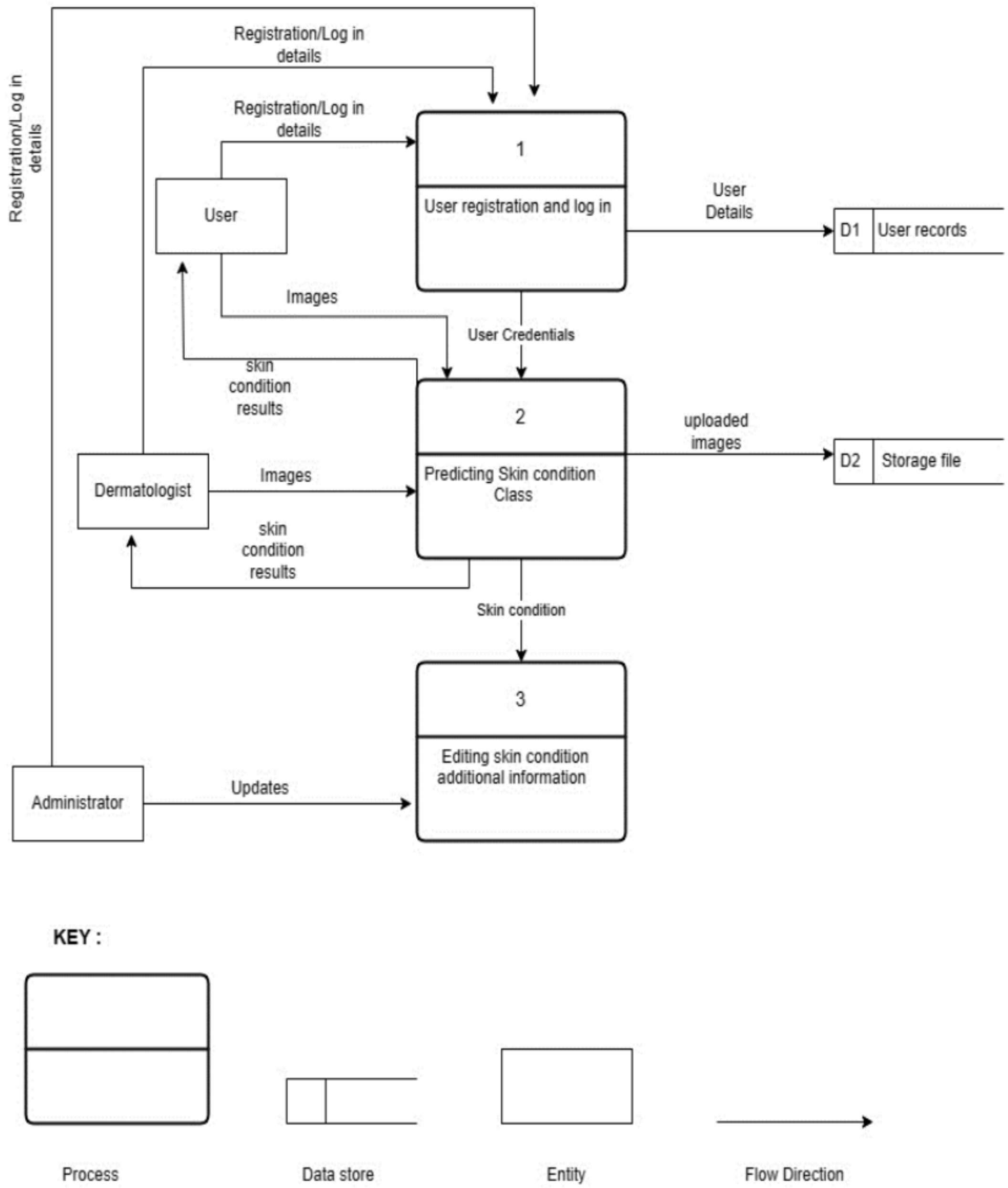


Fig 4.3: Data flow diagram

4.4 Physical design

An application's architecture and system parts that make up the application are referred to as its physical design. It consists of the software and hardware elements that interact to deliver the required functionality. Dermoscopic images obtained with the use of a dermoscope make up the HAM 10000, which is used to train the machine learning model. Dermoscopes are offered as standalone instruments. You can slide a dermoscopic device over your smartphone camera and position it so that it is centred over the lens of the back camera to attach it to your mobile device.

4.5 Database design

Database design is the process of developing an extensive database data model. Finding the data to be saved, setting it up in tables and columns, and specifying the connections between the tables are all part of this process. Creating a database that is effective, simple to use, and capable of supporting the necessary functionality is the goal of database design. A good database should have low data redundancy, data integrity, and data consistency.

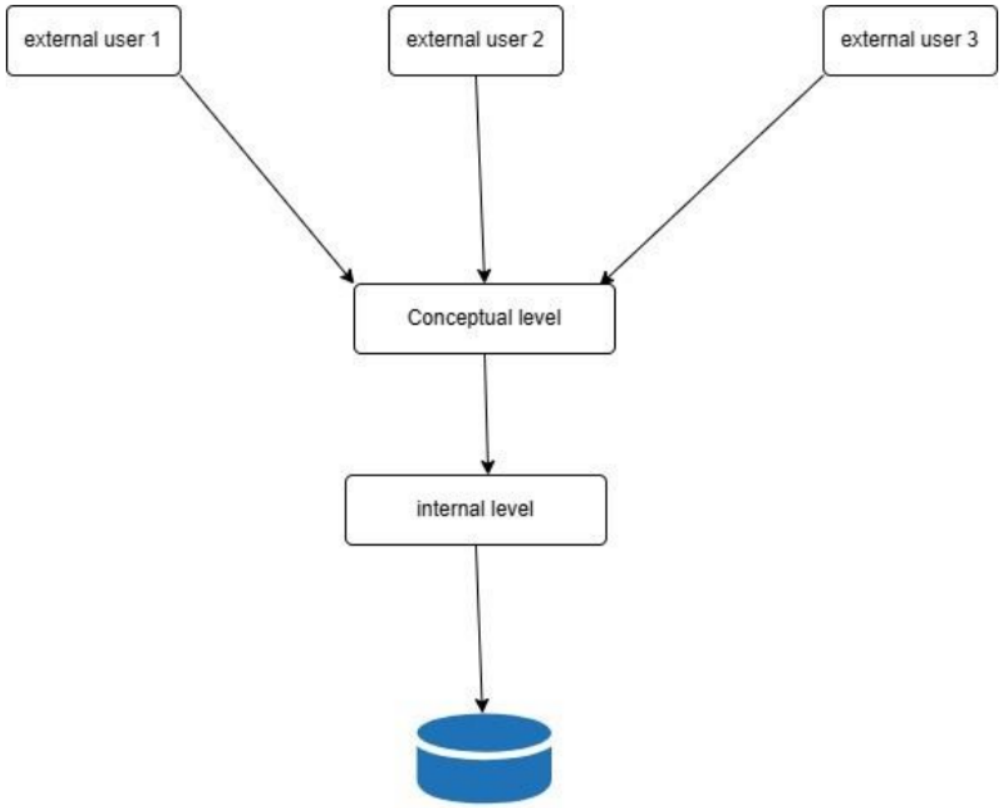


Fig 4.4: Database design

External level

The external level of a database refers to the view of the database that is presented to the end-users or application programs. At this level, the database describes the information that is relevant to the user. At the external level, users are only concerned with the parts of the database that are relevant to their specific tasks or applications.

Conceptual level

The entire logical structure, design, and organization of the data within a database is referred to as its conceptual level. The connections between the data and their arrangement into tables or entities are described at this level. It is concerned with the system's management, storage, and access to data as well as the database's overall structure. The primary goal of the conceptual level is to provide the material in a simple, concise, and usable manner. It offers a framework for arranging data in a way that enables effective information storage and retrieval.

Internal level

The lowest level of abstraction in a database management system is the internal level of a database. It explains how information is kept in the database and accessed by the system. This level describes how data is physically stored, including how it is arranged into files and records. The internal level also describes the methods and data structures that are utilized to retrieve and alter data, as well as how the system accesses and manipulates the data. Anyone working with databases has to understand the internal level because it serves as the basis for all other levels of abstraction in the system.

Entity relationship diagram

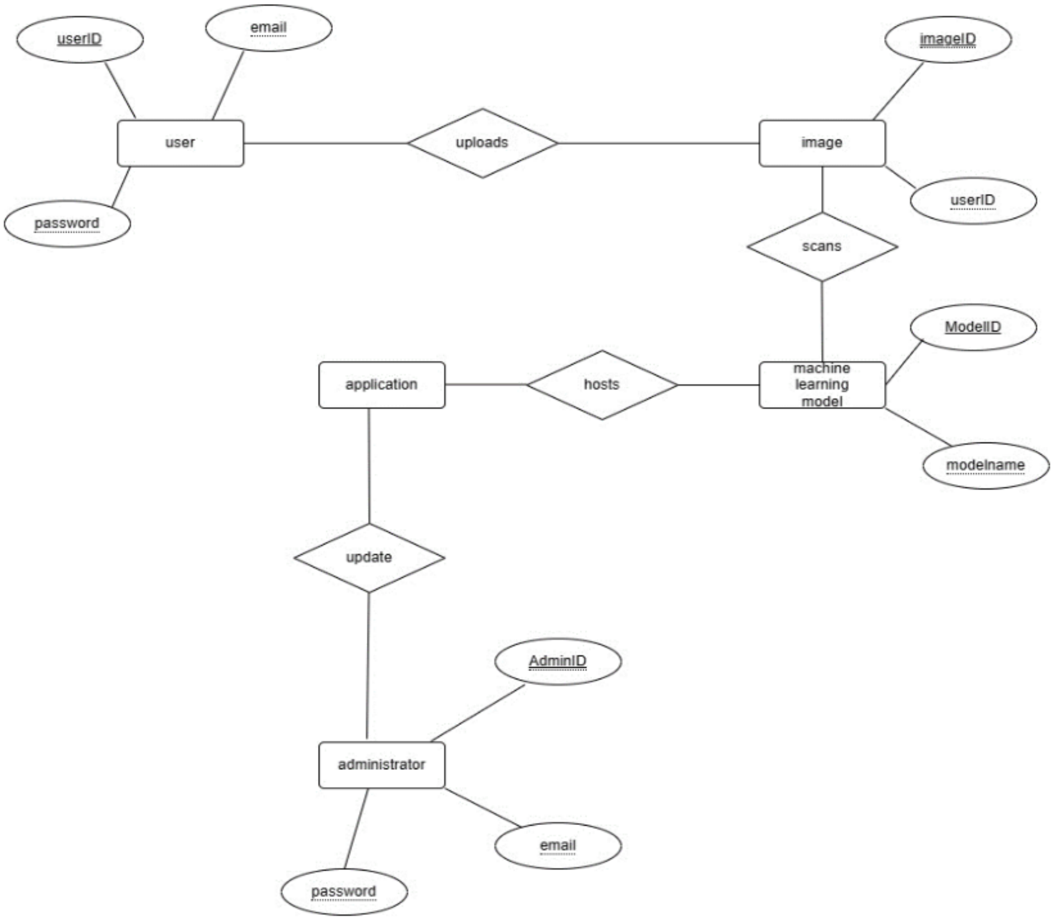


Fig 4.5: Entity relationship diagram

Table 4.1: User table

Column	**Data type**	**Description**
userId	INT	Unique identifier
User_name	VARCHAR	Username for login
password	VARCHAR	Password for login

Table 4.2: Administrator table

Column	**Data type**	**Description**
adminID	INT	Unique identifier
User_name	VARCHAR	Username for login
password	VARCHAR	Password for login

Table 4.3: Skin lesion table

Column	Data type	Description
lesionID	INT	Unique identifier
userID	INT	Foreign key referencing user
Image	BLOB	Uploaded image
Scanned_Result	VARCHAR	Classification result obtained after scanning

4.6 Interface design

The design of various software and hardware interfaces that allow users to engage with computers and other technology is known as user interface design, or UI design. Software development must include interface design since it affects how users interact with a system. While a poorly designed interface can cause frustration and mistakes, one that is well designed can enhance user experience and increase efficiency. The new system is made to be simpler for a user to interact with it and to cut down on the time spent attempting to access the offered services.

4.6.1 Menu design

A menu is a collection of options or guidelines given to a user to carry out a certain action or function. Menus can be displayed in a variety of ways, such as pop-up, drop-down, and context menus. Additionally, menus can aid users in navigating through complicated interfaces and offer an easy way to access a system's various features and functionalities.

4.6.1.1 Main menu

A graphical user interface (GUI) component called Main Menu is used in software programs to give users a list of options and commands to choose from. It is typically the initial screen a user sees when an application is launched. Users may easily and conveniently access all of the software's features and functionalities through the main menu. Users can navigate and locate what they're looking for with ease because of the list of options and instructions' simplicity and organization.

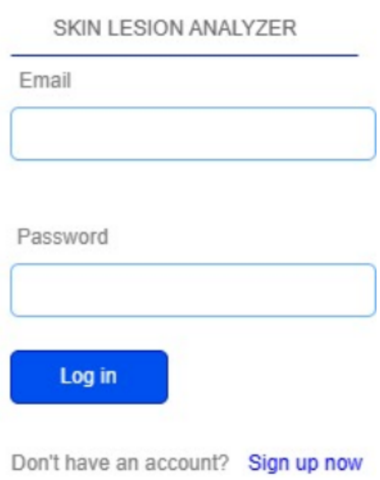


Fig 4.6: Main menu

4.6.1.2 Sub-menu

A sub menu is a secondary menu that is typically displayed as a dropdown menu from a primary menu item. The user can choose from a list of related alternatives to carry out a certain action or move to another area of the website or application. Submenus are frequently used in navigation bars to compactly and efficiently present a large number of options.

Fig 4.7: Sub-menu

4.6.2 Input design

The process of developing a strategy or plan that specifies how information is submitted to an application or system is known as input design. Correct data validation needs to be ensured by careful form design. When incorrect input is identified it should be rejected, and the user must be notified. In this case, the input design is used by the upload image form, the login form, and the registration input form. The system which automatically classifies skin cancer lesions has the following input formats.

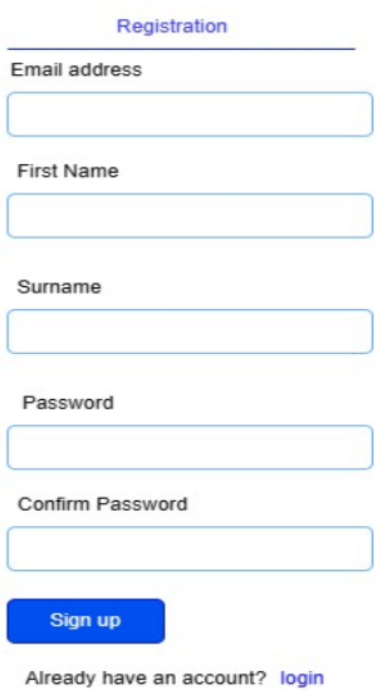


Fig 4.8: Registration form

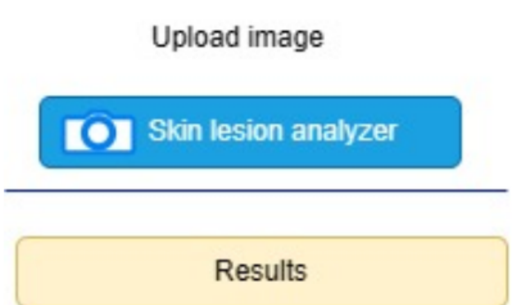


Fig 4.9: Upload image form

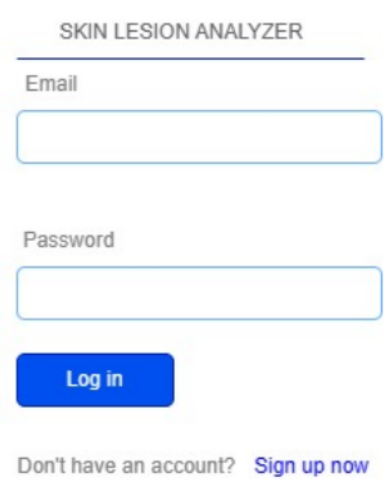


Fig 4.10: Log in page

4.6.3 Output design

After the system's inputs have been received and processed, an output design is used. The process of developing presentable and comprehensible visual representations of data or information is referred to as output design. It entails using a variety of tools and methods to create graphs, tables, charts, and other types of data visualization. In this case, the categorization results and any additional information required are shown using the output design.

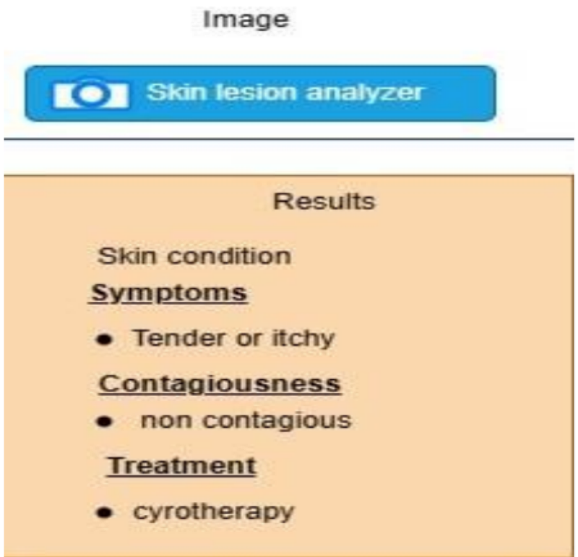


Fig 4.11: Results output page

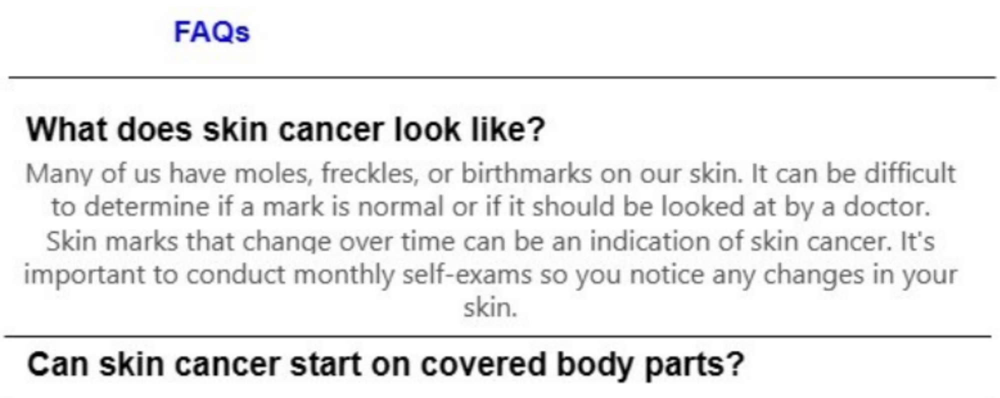


Fig 4.12: FAQs page

Fig 4.13: Contact details page

4.7 Pseudo code

Registration

Is the user logged in?

Redirect the user to "sign up" if they are not.

Notify the user if they don't enter the necessary information.

Inform the user if there is a passwords mismatch if the initial password & the second password entry inputs do not match

Once the necessary forms have been filled, add the user's data that was input into the database.

else, show an error message.

User log in

if username and password are valid

Allow user to access app

else

Show error message

upload image

if image is png or jpg

upload image for analysis

else show error message

Analyze image

Use ML model to classify uploaded image

Display results to user with the correct classification

Display additional information on the lesion classification

update application details

if administrator’s username and password are valid

allow administrator to make changes or Update app

Show success message

else

Show error message

4.8 Security design

Security design is an essential aspect of building any system that deals with sensitive data or information. Security design is to secure data from all dangers and to ensure that unauthorized access to systems or data is avoided. A variety of safeguards are put in place to prevent unwanted access or data breaches as part of security design. Access controls, encryption, firewalls, intrusion detection systems, and other safeguards are among them. Operating rules and procedures for the system must be developed in a number of ways that consider the physical and logical security.

4.8.1 Physical security

In this context, physical security describes the safeguarding of the physical assets that host software programs. These resources include the hardware needed to support the software program, such as servers, data centres, storage devices, and other items. Access controls, like locks and security cameras, as well as environmental controls, and backup power sources, are examples of physical security measures. In addition to these precautions, physical security also includes the rules and guidelines that control how these resources are used, such as security audits and user training programs. The possibility of physical threats to the software applications and the data they contain can be minimized by setting up a thorough physical security approach.

4.8.2 Network security

Network security describes the practice of protecting computer networks from unauthorized access and cyber-attacks. By creating a secure environment that permits allowed access to network resources and data, network security's main objective is to stop network-based attacks. Access control, IDS, firewalls, (VPNs), and encryption are a few techniques used to safeguard application networks. For the network and application to be accessible to only authorized users, access control and authentication are crucial security procedures. To prevent illegal access, it is advised to employ strong passwords, multi-factor authentication, and role-based access control techniques. To identify threats on a network and take appropriate action, an intrusion detection system (IDS) is utilized. Secure communications between devices are made possible via virtual private networks (VPNs) over the internet. Information is transformed into a code through the process of encryption to prevent unwanted access to the data.

4.8.3 Operational security

Operational security describes the process of identifying, analysing, and implementing measures to protect sensitive information and activities from unauthorized access, exploitation, or surveillance. Operational security aims to stop attackers from gathering or assembling data that could be used to compromise the security, integrity, or confidentiality of a patient's medical records. When using the system, users must feel secure. Data encryption, user education, data anonymization, model protection, incidence response, and regular updates and patching are important areas to concentrate on to maintain operational security.

4.9 Conclusion

In conclusion, the design phase is a crucial part of any project as it sets the foundation for the development and implementation stages. During the designing stage, a detailed description of how the proposed system will operate in daily life was provided. There was also a description of each phase of designing, such as database, architectural and application design, illustrating the structure of the system as a whole. The design is generated, reviewed, and improved until it satisfies the project's criteria and goals. The following chapter will highlight testing techniques used throughout the project's development as well as include and explain code snippets. The maintainability and future development of the dermatological application will also be discussed in the following chapter.

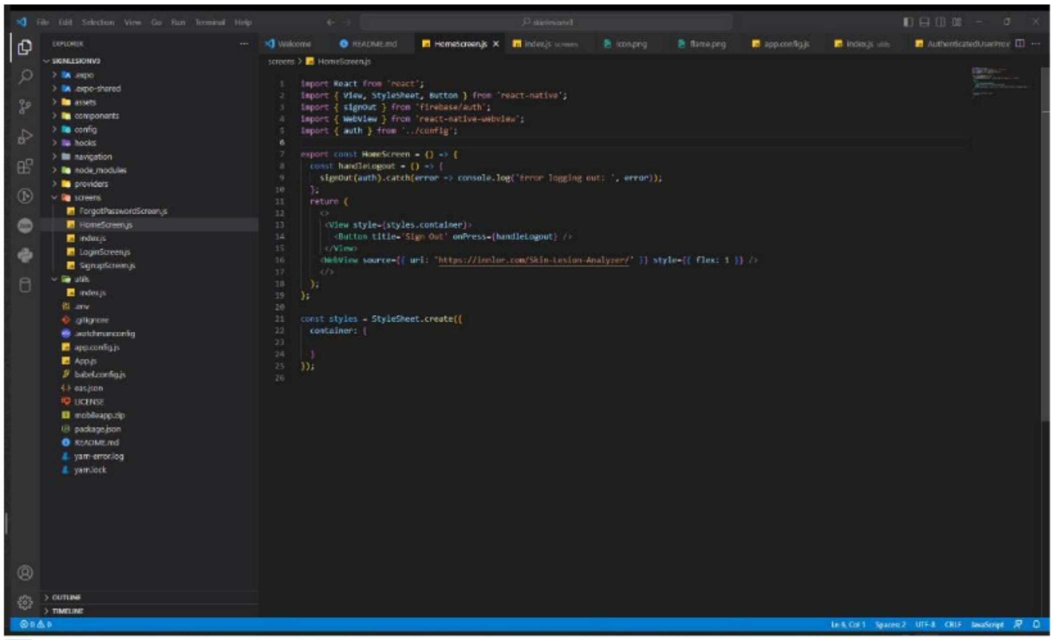
Chapter 5: Implementation Phase

5.1 Introduction

Implementation, in the context of software development, refers to the creation, testing, and deployment of computer code that has been created to address a particular issue or need; in other words, it refers to the process of putting a strategy or solution into practice. The implementation phase, which entails carrying out the strategy that has been devised in the earlier phases, is a vital stage in any project. Metrics such as precision, recall and f1 scores will be employed to provide a quantitative measure of the systems performance The chapter will also include recommendations on how to maintain the system over time and on how to improve the system.

5.2 Coding

The process of developing applications, websites, and other computer programs using an organized set of instructions which a computer can comprehend and carry out is referred to as coding, also known as programming.



28

Fig 5.1: home page

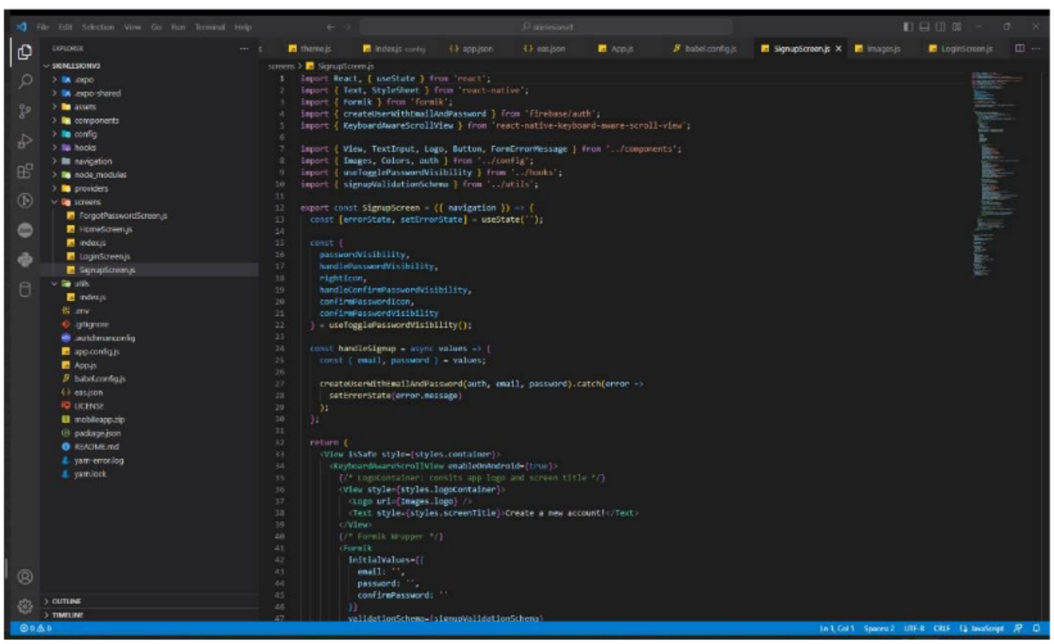


Fig 5.2: Sign up

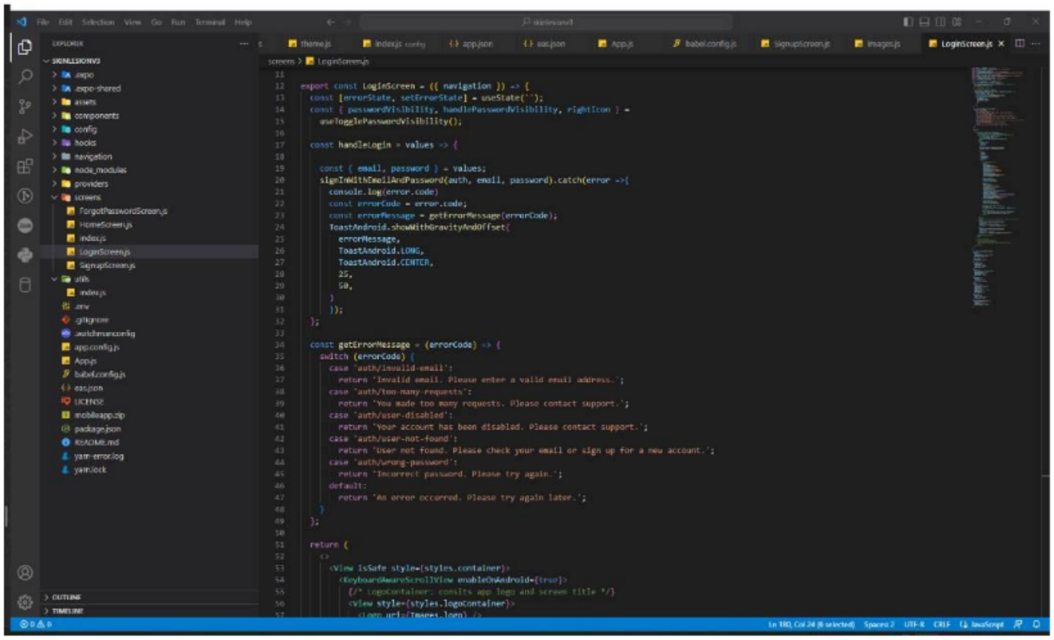


Fig 5.3: Log in

5.3 Testing

Software testing is a means of assessing and verifying that a software application or product performs as intended. Any procedure used to design an application must include testing. Testing has advantages such as preventing bugs and lowering development expenses. Even when a system does not appear to have any obvious problems, testing may still be conducted to enhance its performance. To make sure the application is operating as planned and that it satisfies the demands of its users, testing should be done continuously throughout the development process. Software that is meant for medical use must undergo extensive testing and assessment to verify that it complies with legal standards and is secure and reliable for the use for which it is designed.

5.3.1 Unit testing

The smallest element of an application which can be tested is referred to as a unit. A specific kind of software testing called unit testing involves checking the functionality and proper operation of individual software program units or components in isolation. The log-in unit, sign up unit, and skin lesion classifier unit make up this application.

5.3.2 Integration testing

Software testing includes an integration testing level or phase where software components or modules are joined and tested collectively to gauge their interaction and compliance to functional requirements. Integration testing helps ensure that the software application's various parts or modules are compatible with one another and function together as intended. Before the application is made available to users, integration testing aids in the early detection of defects or issues in software. As issues that are found early are often simpler and cheaper to fix compared to those that are identified later, this may contribute to lowering the amount of time and money required to fix difficulties.

5.3.3 Confusion matrix

An analysis of a machine learning model's performance on a set of test data is summarized by a confusion matrix. It is frequently used to assess how well categorization models work. These models try to identify a categorical classification for each input event. The matrix shows how many genuine positives (TP), true negatives (TN), fake positives (FP), and false negatives (FN) the model generated using the test data. A confusion matrix can be useful for assessing a classifier's performance since it offers a more in-depth look at the model's performance in predicting the various classes. Additionally, it may be used to figure out different performance indicators including F1 scores and accuracy, precision and recall.

Table 5.1: Confusion matrix

True Label	**akiec**	11	3	2	3	5	2	0
	bcc	1	26	0	1	2	0	0
	bkl	6	5	10	2	24	28	0
	df	0	0	0	3	0	3	0
	mel	3	3	0	2	19	11	1

	nv	1	27	1	19	19	681	3
	vasc	0	0	0	0	0	3	8
		akiec	**bcc**	**bkl**	**df**	**mel**	**nv**	**vasc**
Predicted label								

5.3.4 Precision

Precision is the percentage of accurately predicted positive instances (True Positives) out of all anticipated positive instances when testing machine learning models. Typically stated as a percentage as well as a value in decimals between 0 and 1, it provides a gauge of how well the model is able to recognize positive cases. Precision is a crucial statistic in machine learning assessment since it aids in determining the model's dependability and utility. Low precision shows that the model is subject to false positives, while high precision indicates the model is producing accurate predictions.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives + False Positives}}$$

5.3.5 Recall

Recall in machine learning is the proportion of positive cases (True Positives) that were properly predicted out of all positive instances. It is usually stated as a percentage or a decimal between 0 and 1. It is a measurement of how successfully the model can detect all positive cases. Recall is a crucial parameter on machine learning evaluation since it aids in determining how comprehensive and complete the model is. A model with a high recall will be able to recognize every positive prediction, whereas one with a poor recall will be more likely to overlook positive instances.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives + False Negatives}}$$

5.3.6 F1 score

The F1 score is a statistic used in machine learning which incorporates precision and sensitivity into one value. Its value, which ranges from 0 to 1, is the harmonic mean between precision and recall, with larger values indicating greater performance. A high F1 value shows good precision and recall performance for the model, whereas a low F1 score shows poor sensitivity and precision. The F1 score is helpful when it's crucial to strike the right balance between precision and recall. For instance, it is crucial to strike a balance between precision and recall while using a medical diagnosis application because both false positives and false negatives might have major repercussions. In these circumstances, the model's performance can be assessed using the F1 score.

$$\text{F1 score} = \frac{2 \text{ X Precision X Recall}}{\text{Precision + Recall}}$$

Table 5.2: F1 score

Condition	**Precision**	**Recall**	**F1 score**
akiec	0.50	0.42	0.46
bcc	0.41	0.87	0.56
bkl	0.77	0.13	0.22
df	0.10	0.50	0.17
mel	0.28	0.49	0.36
nv	0.94	0.73	0.92
vasc	0.67	0.81	0.70

Fig 5.4: automatic lesion classification

5.3.7 Input validation

The process of making sure the user's input to a software application is valid and adheres to the expected formatting and data type is known as input type validation.

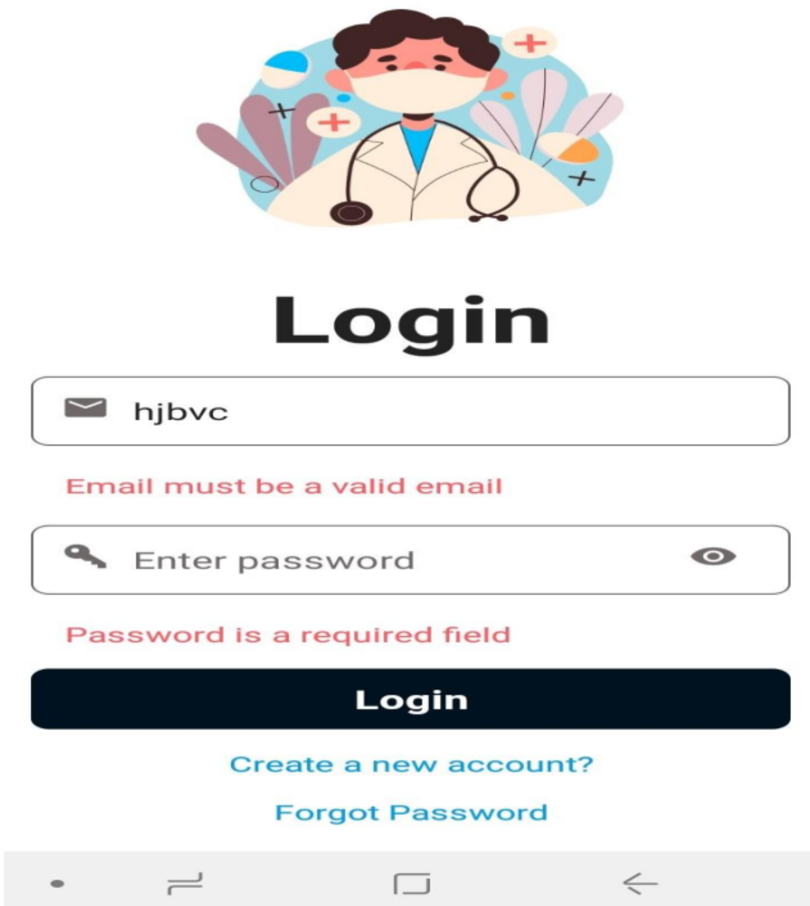


Fig 5.5: validate user input

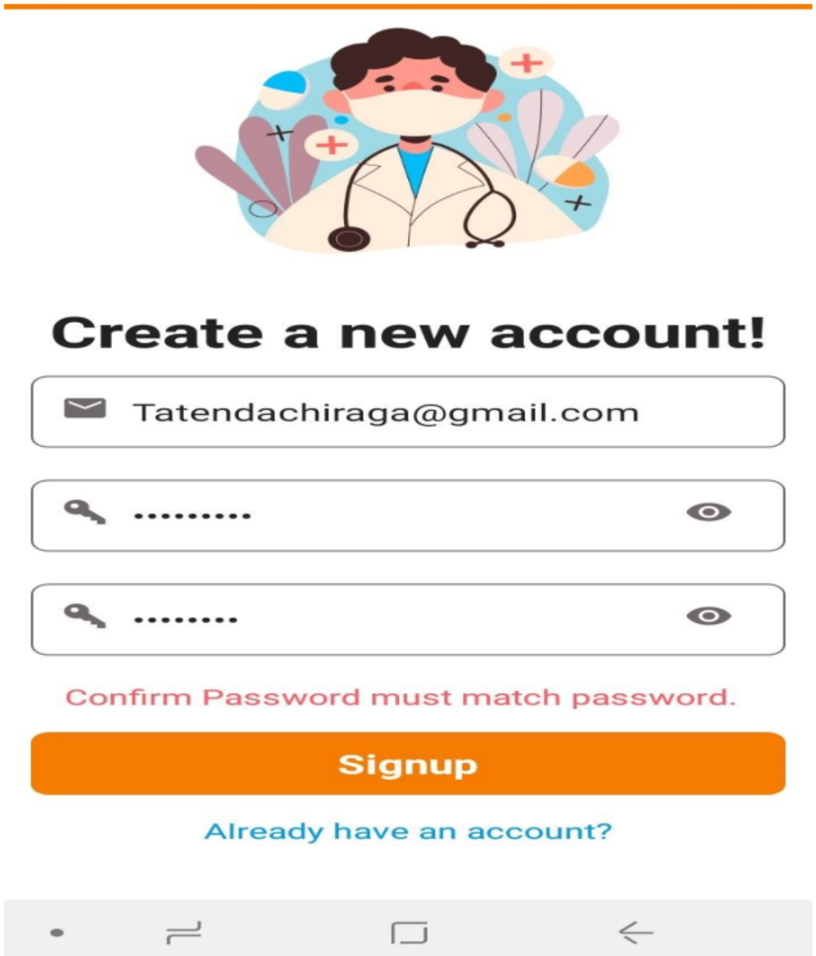


Fig 5.6 validate password matching

5.4 Installation

Installation is the process that involves setting up and deploying a software program onto an electronic device such as a computer in the context of software. The process of creating, testing, and deploying mobile apps is made easier by a combination of services and tools offered by Expo. The TensorFlow.js package can be used to incorporate the model into an Expo project after it has been trained. As a result, the model may be imported and instantly identify the latest images.

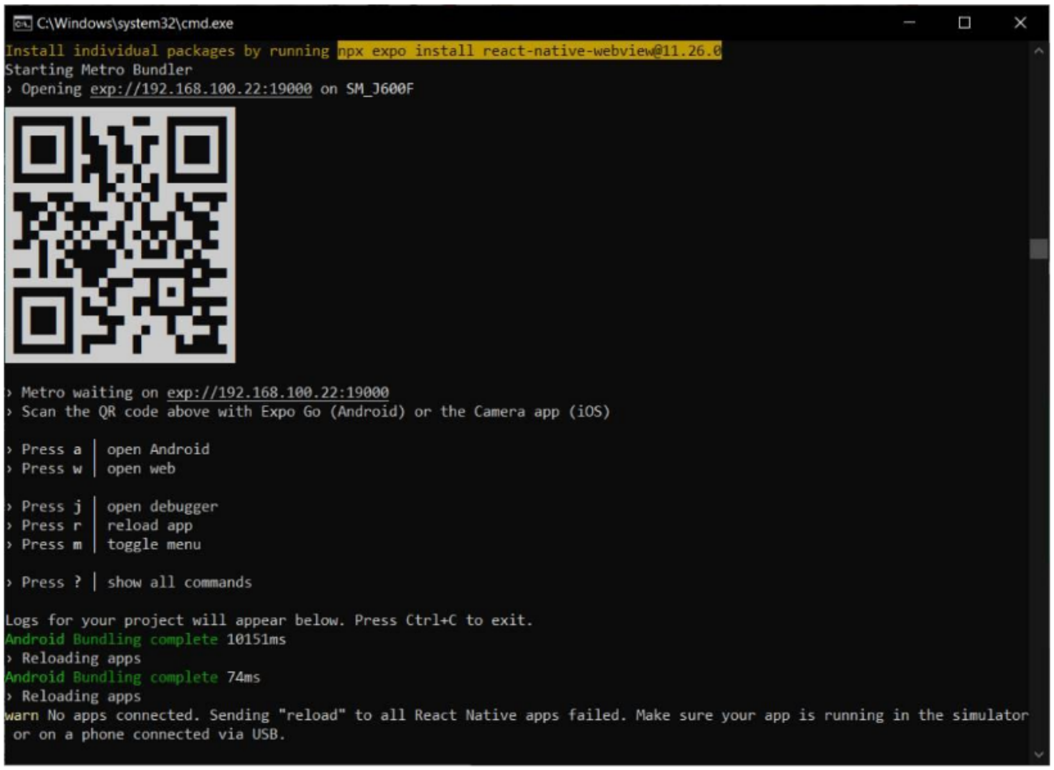


Fig 5.7: expo installation

5.4.1 User training

The practice of showing end users how to utilize a system or piece of software is known as user training. Any application should include user training, but it is especially crucial for complex systems like machine learning because there is a direct correlation between training and system success. Self-instruction, formal classes, and remote/online training are some examples of user training types. Training users on how the application is supposed to be used gives them clear instructions on how to input data, understand the results and navigate the user interface. Additionally, user training clearly outlines the application's objectives, ensuring that users are aware of the program's aims, constraints, and potential applications. The automatic grading system is heavily depended on the image uploaded thus users have to be trained on how they can take clear images which in turn give clear and accurate results.

5.4.2 Data migration

Transferring data from one system or platform to another is known as data migration. Roboflow offers resources for organizing and preparing image collections for machine learning. The dataset must be cleaned and pre-processed to make sure it is correct and comprehensive before being migrated to Roboflow. This can include removing duplicates and adjusting the sizes and formats of the images. The dataset can then be enhanced and annotated after being imported. The data is exported to Google Colab as a zip file after it has been properly augmented and annotated. In Google Colab, the data may be trained, and its performance matrix can be assessed.

5.5 Maintenance

- **Keep iterating and improving**: Continue gathering data, retraining models, and optimizing to improve accuracy over time. A very accurate and reliable solution will be produced through iterative improvements.
- **Address ethical issues**: Machine learning models may bring up ethical issues like security and privacy. Healthcare application providers should make sure that models are created and applied responsibly, keeping in mind patient security and privacy.
- **Regular updates**: Update the app frequently with the most recent iteration of the deep learning algorithm or the training dataset. Over time, this will assist in raising the classification's accuracy. The model could need to be updated once on a monthly basis or once every few years, depending on how quickly skin lesion images change.
- **Continuously assess the model's performance and make improvements**: In order to keep machine learning models current and correct, they must be continually assessed and enhanced. The precision and relevance of a model can be increased by incorporating user feedback and routinely retraining it with fresh data.
- **Document the model and its development**: Keep detailed records of the model's evolution, including information on the data used to train the model, the model's

architecture and hyperparameters, and any updates or modifications that have been made to the model over time. This can make it easier to ensure that the design can be properly reproduced and maintained throughout time.

- **Test the app frequently**: Make sure that the app is frequently checked for bugs and mistakes, and that any problems found are quickly fixed. Machine learning models may contain bias, which could result in unfair or inaccurate predictions. It's crucial to regularly check the model for bias and take action to fix any problems that are found.

5.6 Recommendations for future development

- **The dataset should be expanded**: The models will be better able to learn and become more accurate with more data. To create new training instances, you can also supplement the currently available data. This can be accomplished by gathering the users' scanned images as well as additional real-world data from dermatologists. The model should be trained with a dataset that is diverse enough to represent all types of skin tones.
- **Improve data quality**: To be effective, machine learning models need high-quality data. Healthcare professionals should make sure that the data is accurate, properly annotated, complete, and current. To capture clearer images of the skin lesion, dermatoscopes can be connected to mobile phones.
- **Add additional features**: Include metadata like the patient's age, sex, lesion location, etc. in addition to the images. By identifying patterns that are not readily apparent in images alone, these extra features can provide the model with more context. Subtle patterns might only become apparent when patient metadata is taken into account in addition to visual elements from images. To enhance its capacity for prediction, the model can discover these patterns.
- **Get dermatologists to examine model predictions:** Dermatologists can evaluate some of the model's predictions and offer feedback. Then, to increase performance, retrain the models using the feedback. This interactive loop could improve precision. It offers input

to help the model be retrained. Dermatologists are able to identify incorrect predictions and assign the appropriate labels. The model is able to retrain and enhance its performance using this input. This interactive retraining using human feedback procedure improves the accuracy of models and builds model trust.

- **Explore several other machine learning models:** Try additional models like SVM, random forest, XGBoost, etc. in addition to deep learning models like CNNs. An improved categorization may be provided by an ensemble of several models. Different machine learning models have unique advantages and disadvantages. You can enhance accuracy by finding the model that performs best on the skin lesion classification task by experimenting with numerous different models. The effectiveness of the skin lesion classification model can be increased by investigating novel strategies like transfer learning or self-supervised learning.
- **Ensure transparency and interpretability**: keep in mind that machine learning models might be complex and challenging to understand. Therefore, establishing techniques to comprehend why a model makes particular predictions might aid in increasing the model's credibility and usability.
- **User support**: Users using the application may run into problems or have inquiries about the model and its predictions. User assistance, like a help desk or FAQs, can help guarantee that users can use the application successfully.

5.7 Conclusion

According to the quantitative evaluation of the skin lesions classification model, the model was highly trained on the classes of melanocytic nevi, vascular skin lesions, basal cell carcinoma, and actinic keratosis, and less so on the classes of benign keratosis, dermatofibroma, and melanoma. Since machine learning is a continually changing industry, suggestions on how to maintain and further develop the system are essential to ensuring that the application can stay up with the times and enhance its usability.

FIRST DRAFT

ORIGINALITY REPORT

12%	9%	2%	9%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Midlands State University Student Paper	6%
2	ir.msu.ac.zw:8080 Internet Source	1%
3	nopr.niscair.res.in Internet Source	1%
4	Submitted to University of Nottingham Student Paper	<1%
5	ageconsearch.umn.edu Internet Source	<1%
6	Submitted to The Hong Kong Polytechnic University Student Paper	<1%
7	Submitted to University of Stirling Student Paper	<1%
8	ltu.diva-portal.org Internet Source	<1%
9	"Machine Learning with Health Care Perspective", Springer Science and Business	<1%

Media LLC, 2020
Publication

10	www.dfki.de Internet Source	<1 %
11	m.mu.edu.sa Internet Source	<1 %
12	Submitted to Asia Pacific University College of Technology and Innovation (UCTI) Student Paper	<1 %
13	Submitted to Liverpool John Moores University Student Paper	<1 %
14	www.alliedacademies.org Internet Source	<1 %
15	discovery.ucl.ac.uk Internet Source	<1 %
16	Submitted to University of Surrey Student Paper	<1 %
17	focalplane.biologists.com Internet Source	<1 %
18	Submitted to Shanghai Pinghe School Student Paper	<1 %
19	Submitted to Wigan and Leigh College, Greater Manchester Student Paper	<1 %

20	www.kaggle.com Internet Source	<1 %
21	"Soft Computing: Theories and Applications", Springer Science and Business Media LLC, 2021 Publication	<1 %
22	Submitted to Info Myanmar College Student Paper	<1 %
23	Submitted to University of Portsmouth Student Paper	<1 %
24	research.rug.nl Internet Source	<1 %
25	www.sersc.org Internet Source	<1 %
26	Submitted to Harare Institute of Technology Student Paper	<1 %
27	Jayani Withanawasam, Ehsan Javanmardi, Yanlei Gu, Shunsuke Kamijo. "Modeling Required Driver Attention Level Based On Environmental Risk Factors Using Deep Convolutional Neural Networks", International Journal of Automotive Engineering, 2021 Publication	<1 %
28	ijircce.com Internet Source	<1 %

29	Submitted to Multimedia University Student Paper	<1 %
30	www.coursehero.com Internet Source	<1 %
31	dc.uwm.edu Internet Source	<1 %
32	sd.blackball.lv Internet Source	<1 %
33	"Machine Learning in Medical Imaging", Springer Science and Business Media LLC, 2019 Publication	<1 %
34	en.wikipedia.org Internet Source	<1 %
35	ijarsct.co.in Internet Source	<1 %
36	repository.dl.itc.u-tokyo.ac.jp Internet Source	<1 %
37	researchspace.ukzn.ac.za Internet Source	<1 %
38	www.tutorialaicsip.com Internet Source	<1 %