Skin Cancer Prediction

An Engineering Project in Community Service

**Phase – II Report**

### INTRODUCTION

Skin cancer stands as a pervasive health concern that extends beyond clinical boundaries, affecting millions globally. The challenges inherent in delayed diagnoses and therapeutic complexities underscore the urgent need for a paradigm shift. This report endeavors to explore the transformative potential residing within contemporary technologies, notably machine learning and artificial intelligence, to redefine our comprehension and management of skin cancer.

Driven by empirical observations and substantiated by prevailing challenges, the profound significance of these pioneering tools becomes apparent. They offer a promising avenue for early prediction and precise identification of skin cancer, potentially mitigating mortality rates and alleviating the burden on both individuals and society.

The focal intent of this exploration is to underscore that these advancements transcend conceptual realms; they embody pragmatic solutions deeply embedded in empirical narratives. They bestow upon individuals bespoke strategies, fortifying their defenses against skin cancer risks. This discourse doesn’t solely center on technological innovations; rather, it charts a course toward leveraging technology to tangibly transform the lives of those affected by skin cancer. It accentuates the pivotal role of early detection protocols and individualized care in this pursuit.

Motivation:

Skin cancer represents a significant global health concern, with delayed diagnoses and complex treatments posing substantial challenges. Consequently, there's an urgent need for innovative approaches to redefine how we understand and manage this condition. By harnessing the power of modern technologies like machine learning and artificial intelligence, we have the potential to revolutionize skin cancer detection and treatment, ultimately reducing mortality rates and easing the burden on individuals and society.

1. EXISTING WORK / LITERATURE REVIEW

## Table 2 : Literature Review/Pre-existing work

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sr. No. | Reference | Year of Publication | Title | Description | Authors |
| 1. | Javaid, A., Sadiq, M., Akram, F. (2021) Skin Cancer Classification Using Image Processing and Machine Learning. In: International Bhurban Conference on Applied Sciences and Technologies.  Islamabad, Pakistan.  pp. 439-444. | 2021 | Skin Cancer Classification Using Image Processing and Machine Learning. | * Classification accuracies   achieved on the ISIC-ISBI2016  dataset using SVM, Quadratic Discriminant, and Random Forest classifiers 85.50%, 88.17%,  90.84%,  respectively.   * Among are and   114 malignant  guesses, 89.7% are accurate and 5.3% are  erroneous. | Javaid, A., Sadiq, M., Akram, F. |
| 2. | P. Dubal, S. Bhatt, C. Joglekar and S. Patil, "Skin cancer detection and classification," 2017 6th International Conference on Electrical Engineering and Informatics, Langkawi, Malaysia, 2017, pp. 1-6, DOI:  10.1109/ICEEI.2017.8 | 2017 | Skin cancer detection and classification | * Thedataset was divided into training, testing, and validation   sets, each consisting of 80%, 10%, and  10% of the dataset, respectively.   * Classified 463 | P. Dubal, S. Bhatt, C. Joglekar and S. Patil |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 312419. |  |  | images into their respective categories, with an overall accuracy rate of 76.9%,   * Algorithm on the   basis of four features i.e. asymmetry, border, color, and diameter |  |
| 3. | Emara, “A Modifier Inception\_v4 for  imbalanced skin Cancer Classification”, 2019 14th International Conference on Computer Engineering and Systems , pp. 28-33, December 2019. | 2019 | A Modifier Inception\_v4 for imbalanced skin Cancer Classification | * Thedataset was divided into two parts, 90% and   10% used for training and validation, respectively.   * Thedataset   images were resized to 299 ×  299 to be compatible with the Inception-v4 architecture. | Emara |
| 4. | Y Jusman,”  Performance of Multi-Layer Perceptron and Deep Neural Networks in Skin Cancer Classification”, IEEE Transactions, vol 10, pp. 118198 | 2022 | Performance of Multi-Layer Perceptron and Deep Neural Networks in Skin Cancer Classification | * Trained Multi-layer Perceptron,   custom CNN, and VGG-16 on HAM10000  dataset for skin cancer classification. | Y Jusman |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | -118212, 2022. |  |  | VGG-16 exhibits superior accuracy;   * VGG-16 and   custom CNN are faster than Multi-layer Perceptron. |  |
| 5. | R Raja Subramanian, “Skin Cancer Classification using CNN”, IEEE  Conference, 2021 11th International  Conference on Cloud Computing, Data Science & Engineering 10.1109/Confluence51 648.2021.9377155,  2021. | 2021 | Skin Cancer Classification using CNN | * Various models and   methodologies were implemented;  some yielded better accuracy.   * CNN trained   with augmented images, utilizing a dataset of nearly 23,000 images, produced   * superior results.   Achieved an accuracy of more than 80% with the HAM10000  dataset. | R Raja  Subramania n |
| 6. | Harikrishna, “Skin Cancer Classification using Transfer Learning IEEE International  Conference on Advent Trends in | 2020 | Skin Cancer Classification using Transfer Learning | * Authors prioritize accuracy,   precision, and  recall in  classification evaluation. | Harikrishna |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Multidisciplinary Research and Innovation  (ICATMRI-2020) |  |  | * After removing duplicate images, * An 83:17   validation split is created.  ResNet50, with 90% accuracy, emerges as the final model choice |  |
| 7. | Nourabuared, “Skin Cancer Based on VGG19 and Transfer Learning”, In 3rd International  Conference on Signal Processing and Information Security IEEE Conference, pp 1-4,  DOI-https://doi.org/10. 1109/ICSPIS51252.202  0.93401432021. | 2021 | Skin Cancer Based on VGG19 and Transfer Learning | * Study focuses on segmenting and categorizing skin lesion images for remote cancer identification, avoiding hospital visits. * Combines   MeanShift segmentation with kNN,  decision trees,  and SVM classification, achieving an 86.5% accuracy for diagnosing skin cancer | Nourabuare d |
| 8. | Enakshi Jana, “Research on Skin cancer Cell Detection using Image  Processing”, 2017 | 2017 | Research on Skin cancer Cell Detection using Image Processing | * BNN classifier achieves 89.9% accuracy, while AANN achieves 80.8% accuracy. | Enakshi Jana |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | IEEE International Conference on Computational  Intelligence and Computing Research, DOI: 10.1109/ICCIC.2017.8  524554. pp. 1-8, 2017. |  |  | * SVM and AdaBoost yield the best results for skin cancer detection; a   description of Melanoma aids in normal and abnormal skin cell classification. |  |
| 9. | N. J. Dhinagar and M. Celenk, "Analysis of regularity in skin pigmentation and vascularity by an optimized feature space for early cancer classification," 2014 7th International Conference on Biomedical Engineering and Informatics, Dalian, China, 2014, pp.  709-713, DOI:  10.1109/BMEI.2014.70  02865. | 2014 | Analysis of regularity in skin pigmentation and vascularity by an optimized feature space for early cancer classification | * The selected features form a powerful means of discriminating between the three classes of surface-scanne d skin samples * Clustering of   the samples with this feature space results in a classificatio n accuracy of 87.3% | N. J.  Dhinagar and M. Celenk |
| 10. | P. Sedigh, R. Sadeghian and M. T. Masouleh, "Generating Synthetic Medical Images by Using GAN to Improve CNN | 2019 | Generating  Synthetic Medical Images by Using GAN to Improve CNN  Performance in | * Has employed the GAN algorithm to generate   synthetic skin cancer | P. Sedigh, R.  Sadeghian and M. T. Masouleh |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Performance in Skin Cancer Classification," 2019 7th International Conference on  Robotics and Mechatronics, Tehran, Iran, 2019, pp.  497-502, DOI:  10.1109/ICRoM48714.  2019.9071823. |  | Skin Cancer Classification | medical  images based on primary database, which was collected from the International Skin Imaging. |  |
| 11. | Dwaipayan Choudhury, “Texture and Color Feature Based WLS Framework Aided Skin Cancer Classification using MSVM and ELM”, 2015 Annual IEEE India  Conference, December 2015. | 2015 | Texture and Color Feature Based WLS Framework Aided Skin Cancer Classification using MSVM and ELM | * DermNet NZ contains 359 images categorized into four classes:   Squamous cell  carcinoma,B asal cell carcinoma,   * Melanoma, and   Actinic keratosis, each with a  resolution of 150\*112.   * These images   depict various parts of the human body's skin. | Dwaipayan choudhury |
| 12. | A. Masood and A. Al-Jumaily, "Semi | 2017 | Semi advised learning and | * Using both   labeled and | A. Masood and A. |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | advised learning and classification algorithm for partially labeled skin cancer data analysis," 2017 12th International  Conference on Intelligent Systems and Knowledge Engineering , Nanjing, China, 2017, pp. 1-4, DOI: 10.1109/ISKE.2017.82  58767. |  | classification algorithm for partially labeled skin cancer data analysis | unlabeled data made the model better at  understanding new images.   * It got around   91.6% accuracy for skin  pictures and  86.5% for tissue samples. | Al-Jumaily |
| 13. | N. C. Lynn and Z. M. Kyu, "Segmentation and Classification of Skin Cancer Melanoma from Skin Lesion Images," 2017 18th International  Conference on Parallel and Distributed Computing,  Applications and Technologies, Taipei, Taiwan, 2017, pp.  117-122, DOI:  10.1109/PDCAT.2017.  00028. | 2017 | Segmentation and Classification of Skin Cancer Melanoma from Skin Lesion Images | * Research   involvements and classifying skin lesion images to aid patients in identifying skin cancer remotely.   * Employing   MeanShift for segmentation and kNN,  decision tree, and SVM for classification yielded a  promising 78.2% accuracy. | N. C. Lynn and Z. M. Kyu |
| 14. | Chee Jen Ngeh, “Deep Learning on Edge | 2020 | Deep Learning on Edge Device for | * Picks the model with the best | Chee Jen Ngeh |

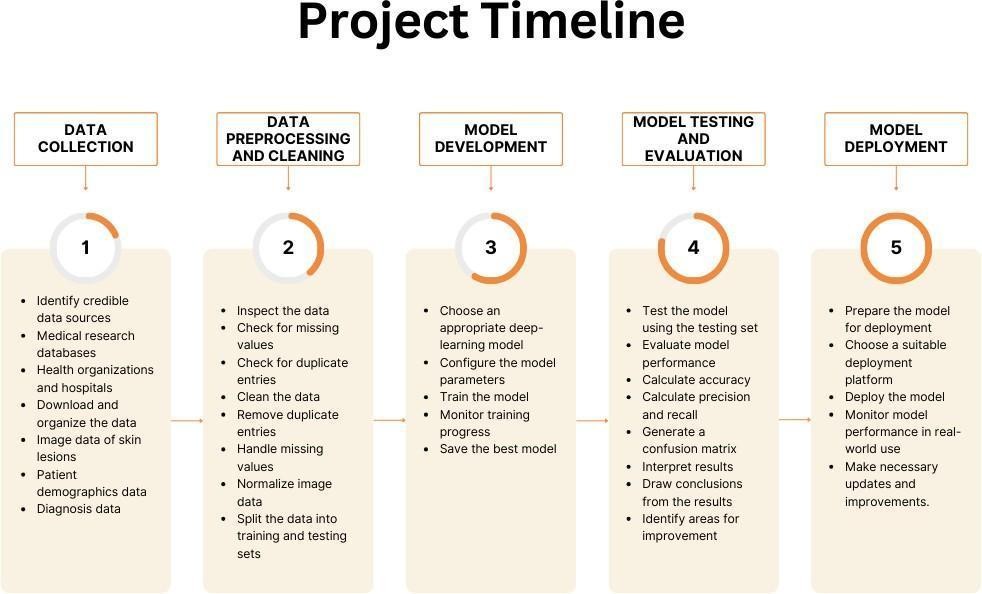
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Device for Early Prescreening of Skin Cancers in Rural Communities, 2020 IEEE Global Humanitarian Technology Conference, DOI:10.1109/GHTC46 280.2020.9342911,  October 2020. |  | Early Prescreening of Skin Cancers in Rural  Communities | performance out of all trials, MobileNetV2, which reaches 72% accuracy,  88% top-2  accuracy, and  94% top-3 accuracy |  |
| 15. | S. R. Guha and S.  M. Rafizul Haque, "Convolutional  Neural Network Based Skin Lesion Analysis for Classifying  Melanoma,"  2019 International  Conference on Sustainable  Technologies  for Industry 4.0 (STI), Dhaka, Bangladesh, 2019,  pp. 1-5, doi: 10.1109/STI47673.2  019.9067979. | 2019 | Convolutional Neural Network Based Skin Lesion  Analysis for Classifying  Melanoma | * 1137 images were utilized for training, while   197 were  designated for testing.   * VGG16 transfer   learning showed 11.65%  accuracy enhancement compared to the base CNN model,   * Achieving   91.07%  accuracy. | S. R.  Guha and  S. M.  Rafizul Haque |

Objective

This aims to explore the transformative capabilities of machine learning and artificial intelligence in addressing the challenges posed by skin cancer. By examining empirical observations and prevailing issues, we seek to highlight the profound significance of these technologies in enabling early prediction and precise identification of skin cancer. Additionally, we aim to emphasize that these advancements are not merely theoretical concepts but practical solutions grounded in real-world scenarios. Our objective is to illustrate how these technologies can provide personalized strategies to empower individuals in their fight against skin cancer, emphasizing the importance of early detection and tailored care protocols.

### ROADMAP

* 1. *Data Collection*
     + Identify credible data sources
     + Medical research databases
     + Health organizations and hospitals
     + Download and organize the data
     + Image data of skin lesions
     + Patient demographics data
     + Diagnosis data
  2. *Data Preprocessing and Cleaning*
     + Inspect the data
     + Check for missing values
     + Check for duplicate entries
     + Clean the data
     + Remove duplicate entries
     + Handle missing values
     + Normalize image data
     + Split the data into training and testing sets
  3. *Model Development*
     + Choose an appropriate deep-learning model.
     + Configure the model parameters
     + Train the model
     + Monitor training progress
     + Save the best model
  4. *Model Testing and Evaluation*
     + Test the model using the testing set
     + Evaluate model performance
     + Calculate accuracy
     + Calculate precision and recall
     + Generate a confusion matrix
     + Interpret results
     + Draw conclusions from the results
     + Identify areas for improvement
  5. *Deployment*
     + Prepare the model for deployment
     + Choose a suitable deployment platform
     + Deploy the model
     + Monitor model performance in real-world use
     + Make necessary updates and improvements.



## Figure3.1 Project Timeline

### WORKING PRINCIPLE

The Skin Cancer Prediction Using Deep Learning Techniques system starts by gathering a vast collection of skin lesion images from Kaggle. These images undergo preparation steps, including resizing and normalization, to

ensure consistency in analysis. The dataset is then divided into training and testing sets. Convolutional Neural Networks (CNNs) are utilized for their proficiency in recognizing patterns in images. CNN learns to identify

signs of skin cancer through layers that scrutinize different aspects of the image. Special mathematical functions are incorporated to enhance the model's ability to grasp intricate patterns. During training, the model's performance is evaluated based on accuracy and learning progress. Once trained, the model is tested on new images to assess its predictive capability. If successful, the model is saved in a deployable format for potential real-world applications, such as aiding in skin cancer diagnosis. This system provides a valuable tool for early detection of skin cancer, potentially improving patient outcomes.

### IMPLEMENTATION

*Dataset:*

In the initial phase of developing the Skin Cancer Prediction Using Deep Learning Techniques system, our focus was on obtaining the input dataset. The process of data collection marks the first substantial step in the development of a machine learning model, as the quality and quantity of data directly influence the model's performance. The dataset comprises 10,015 images and can be accessed using the following

URL from Kaggle: <https://www.kaggle.com/datasets/jayaprakashpondy/skin-cancer-dataset>

*Importing the necessary libraries:*

For this project, the Python programming language will be utilized. Initially, essential libraries will be imported, including Keras for constructing the primary model, sci-kit-learn for dividing the training and testing data, PIL for converting images into arrays of numbers, along with other fundamental libraries such as pandas, numpy, matplotlib, and TensorFlow.

*Retrieving the images:*

In this module, the objective is to extract images from the dataset and transform them into a suitable format for training and testing the model. This process includes tasks such as reading the images, resizing them, and normalizing the pixel values. The images and their associated labels will be retrieved, ensuring uniformity by resizing all images to a standard size of (28, 28) to facilitate recognition. Subsequently, the images will be converted into numpy arrays. This module involves the loading and extraction of preprocessed images along with their respective class labels from the dataset. These images will serve as input for training and evaluating the deep learning model.

*Splitting the dataset:*

In this section, the image dataset undergoes partitioning into training and testing sets, with 80% allocated to training and 20% to testing. This division is crucial for training the model on a subset of the data, validating its performance, and assessing its accuracy on unseen data.

To ensure effective training and evaluation of the model, the dataset is divided into training, validation, and testing subsets. This ensures that the model is trained, validated, and tested on separate portions of the dataset, enabling a comprehensive assessment of its generalization capabilities.

*Model Development:*

The utilization of convolutional neural networks (CNNs) has proven highly effective in image recognition tasks. The fundamental aspect that sets CNN apart from traditional neural networks is the convolution operation.

When an image is inputted, CNN repeatedly scans it to detect specific features. This scanning, or convolution, is controlled by two main parameters: stride and padding type. The initial convolution process generates a series of new frames, each containing information about a particular feature's presence in the scanned image. The resulting frame highlights areas where features are prominent with higher values, while areas lacking such features have lower values. This process iterates a predetermined number of times, utilizing a classic CNN model with only two convolution layers in this project.

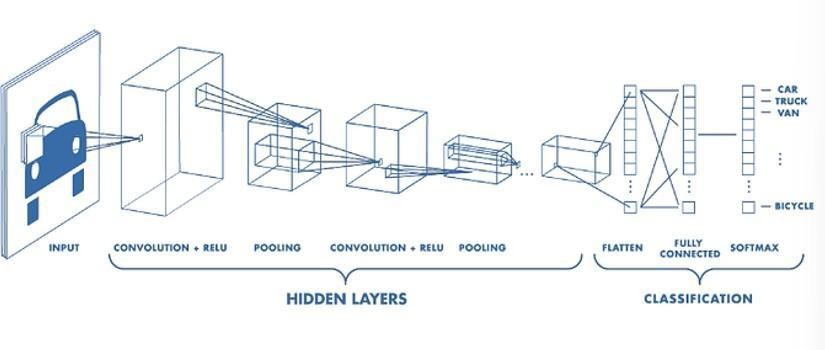
As successive layers are convolved, higher-level features are sought, mirroring human perception. In the context of face recognition, for instance, different CNN layers target distinct features. Initially, if constructing the CNN from scratch, the sought features are random. Through the training process, neuron weights are adjusted, gradually enabling the CNN to identify features crucial for achieving the predefined objective, such as successful image recognition from the training set.

Pooling operations, or sub-sampling, are conducted between these described layers to reduce frame dimensions. Additionally, a nonlinear function, commonly referred to as ReLU, is applied to each resulting frame after convolution to introduce non-linearity to the model.

Towards the end of the network, fully connected layers are employed. The final set of frames derived from convolution operations is flattened into a one-dimensional neuron vector. Subsequently, a standard, fully-connected neural network is applied. For classification tasks, a softmax layer is utilized to transform the model's outputs into probabilities corresponding to each class for accurate classification.

*CNN model Architecture:*

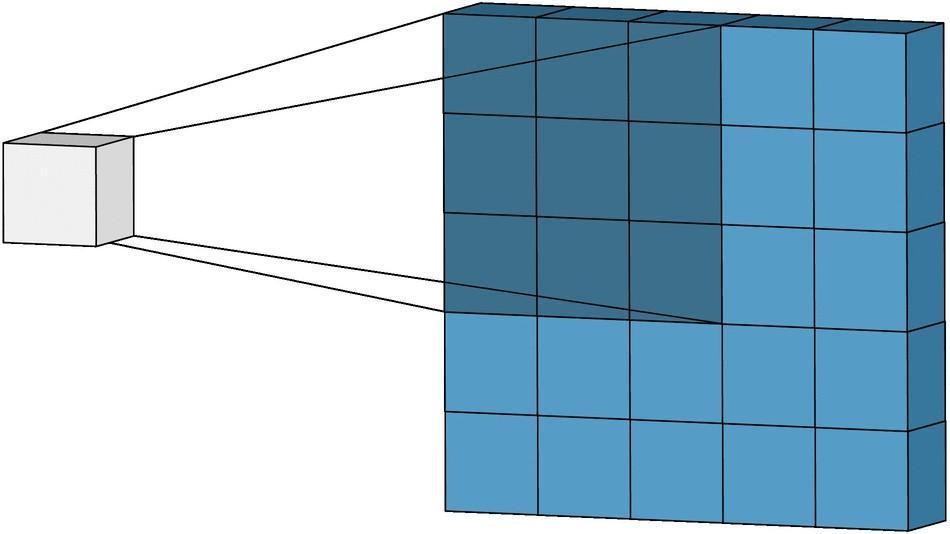
Convolutional Neural Networks (CNNs) follow a specific architecture consisting of three fundamental layers: The convolutional layer, the pooling layer, and the Fully connected layer.



## Figure 5.1 : CNN Architecture

*Convolution Layer*

The heart of the CNN architecture lies the convolutional layer, which serves as the core building block. This layer is responsible for the bulk of the network's computational load. It operates by performing dot product operations between learnable parameters, known as kernels, and restricted portions of the input image, generating activation maps. These activation maps provide a two-dimensional representation of the image, capturing essential features and patterns.



## Figure 5 .2 : Kernel Extraction

The motivation behind the convolutional layer stems from three key ideas: sparse interaction, parameter sharing, and equivariant representation. Sparse interaction is achieved by employing smaller kernels compared to the input image, reducing the number of parameters needed and enhancing the model's memory efficiency. Parameter sharing ensures that the same set of weights is used across different spatial positions, promoting statistical efficiency and reducing redundancy. Additionally, due to parameter sharing, CNN layers exhibit equivariance to translation, meaning that if the input is modified, the output changes in a corresponding manner.

*Pooling layer*

Following the convolutional layer is the pooling layer, which plays a crucial role in reducing the spatial size of the representation. This layer operates by summarizing nearby outputs through various pooling functions, such as max pooling, which reports the maximum output from the neighborhood. By reducing the spatial size, the pooling layer decreases the computational load and the number of weights, contributing to overall efficiency.

*Fully Connected Layer*

The final layer in the CNN architecture is the fully connected layer, where neurons have complete connectivity with neurons in the preceding and succeeding layers. This layer facilitates the mapping between the input and output, allowing for comprehensive representation mapping.

Moreover, non-linearity layers are often integrated after the convolutional layer to introduce non-linearity to the activation maps. Popular non-linear activation functions include Sigmoid, Tanh, and Rectified Linear Unit (ReLU). While Sigmoid and Tanh functions suffer from limitations such as vanishing gradients, ReLU has gained popularity due to its reliability and acceleration of convergence, albeit with potential fragility during training.

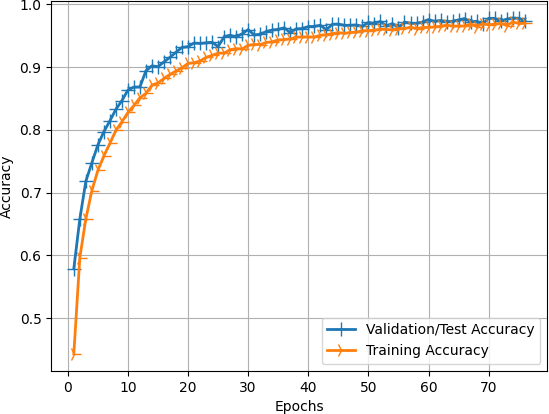
In summary, the architecture of a CNN is meticulously designed to leverage convolution, pooling, fully connected, and non-linearity layers, each contributing uniquely to the network's effectiveness in tasks such as image recognition and classification.

*Apply the model and plot the graphs for accuracy and loss:*

After constructing the model, it will undergo assessment on the validation set to gauge its accuracy and loss. This evaluation process involves plotting accuracy and loss metrics against the number of epochs to visualize the model's performance. Subsequently, we will compile the model and initiate it using the fit function, with a predetermined epoch count of 100. Following this, graphs depicting accuracy and loss will be generated. The obtained average training accuracy stands at 96.00%.

*Accuracy on test set:*

Once the model undergoes training and evaluation on the validation set, its accuracy will be gauged using the test set. This accuracy serves as a crucial metric for assessing the model's performance. Subsequently, a test set accuracy of 97.00% was achieved.



## Figure 5.3: Training and Validation Accuracy over Epochs

*Saving the Trained Model:*

Once reached a point of confidence in your trained and tested model, transitioning it into a production-ready environment requires careful steps. The initial action involves saving the model into a file format suitable for production use, such as .h5 or .pkl, utilizing a library like Pickle for this purpose.

### SYSTEM ANALYSIS

The current system utilizes the RESNET (Residual Neural Network) architecture for skin cancer prediction, renowned for its exceptional image recognition capabilities. Its primary objective was to accurately classify dermatoscopic images into three distinct categories: Melanoma, Basal cell Carcinoma, and Squamous cell skin cancer.

RESNET, a type of convolutional neural network (CNN), integrates skip connections or shortcuts to effectively learn from both shallow and deep layers. This design feature addresses the vanishing gradient problem common in deep neural networks, rendering RESNET particularly suitable for complex image analysis tasks such as skin cancer classification.

Following extensive training on a diverse dataset of dermatoscopic images, the system achieved a notable accuracy of 82.87%. This level of accuracy underscores its capacity to differentiate between the three types of skin cancer with a satisfactory success rate. The classification process involved several stages. Initially,

dermatoscopic images underwent preprocessing to enhance quality and standardize data for the model. Subsequently, the RESNET architecture underwent training on the preprocessed images, learning to extract relevant features and patterns distinguishing the types of skin cancer.

Evaluation of system performance utilized a separate test dataset, distinct from the training data. The trained RESNET model made predictions on this test dataset, resulting in an accuracy of 82.87%, indicating the proportion of correct predictions relative to the total test set samples.While the system's success in achieving a reasonably high accuracy rate underscores the potential of deep learning in aiding medical professionals with early skin cancer detection, it is crucial to emphasize that no diagnostic system is infallible. Human expertise remains vital in final diagnosis and treatment decisions.

Despite encouraging outcomes, ongoing research in medical image analysis continues to refine existing models, explore ensemble methods, and integrate cutting-edge deep learning architectures to further improve accuracy and robustness. Continuous enhancement of skin cancer prediction systems holds promise for improving patient outcomes and alleviating the global burden of skin cancer.

### DISADVANTAGES OF THE EXISTING SYSTEM:

* Challenges in Accuracy for Critical Cases: While achieving an 82.87% accuracy is commendable, it implies that the system may still misclassify a significant number of cases. In critical scenarios where timely and accurate diagnoses are paramount, this level of accuracy could result in delays in administering appropriate treatment to individuals with skin cancer.
* Sensitivity to Image Quality: Deep learning systems, such as the RESNET model, can exhibit sensitivity to variations in image quality, including factors like lighting conditions, resolution, and the presence of noise. Suboptimal image quality or significant deviations from the training data may hinder the system's performance, potentially leading to reduced accuracy.
* Limited Interpretability: Understanding deep learning models like RESNET can be challenging due to their lack of transparency in explaining prediction mechanisms. This lack of clarity poses concerns, particularly in critical medical contexts where the rationale behind decision-making is crucial.
* Data Imbalance Issues: In some cases, there may be insufficient data available for all types of skin cancer, resulting in imbalanced datasets. This imbalance can introduce bias and diminish accuracy, particularly for less prevalent cancer types.
* Resource Intensiveness: The development and training of deep learning models like RESNET require substantial computational resources, such as powerful GPUs or TPUs. This demand for resources can entail significant setup and operational costs, making it challenging to deploy and maintain the system.

### PROPOSED SYSTEM:

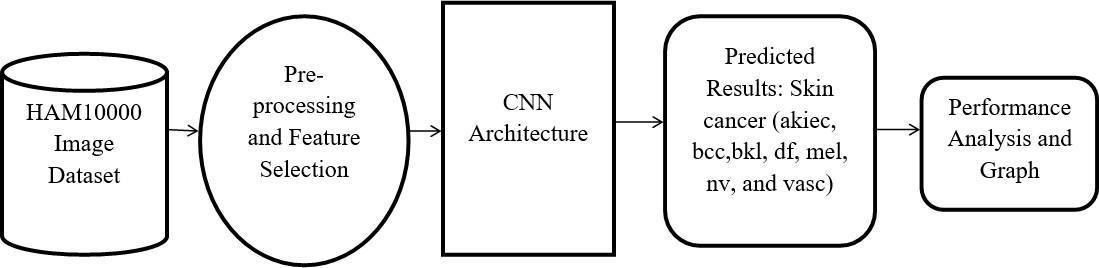
* + Deep Learning Techniques: The proposed system leverages advanced deep learning methods to enhance skin cancer prediction, with a specific focus on analyzing dermatoscopic images.
  + Convolutional Neural Networks (CNNs): The system utilizes CNNs, a type of deep learning architecture, along with tailored enhancements to improve accuracy, efficiency, and overall performance in classifying skin cancer.
  + Cutting-edge Architecture: At the heart of the system lies a state-of-the-art deep-learning architecture, designed to effectively extract intricate patterns from dermatoscopic images and identify critical features associated with various types of skin cancer.
  + Data Augmentation: The system employs data augmentation techniques, such as rotation, scaling, flipping, and cropping, to diversify the dataset and reduce the risk of overfitting, thereby enhancing model generalization.
  + Addressing Class Imbalance: To tackle class imbalance within the dataset, the system implements strategies like oversampling, undersampling, or class weighting, ensuring fair representation of different skin cancer types and improving prediction accuracy.

### ADVANTAGES OF PROPOSED SYSTEM:

* + Enhanced Accuracy: Through the utilization of advanced deep learning architectures and data augmentation techniques, the proposed system attains heightened accuracy in skin cancer classification. This capability stems from its ability to discern intricate patterns from dermatoscopic images, resulting in more precise and dependable predictions.
  + Improved Generalization: The system employs transfer learning and data augmentation to enhance generalization to unseen data, ensuring proficient performance across diverse datasets and handling variations in image quality, lighting, and patient demographics effectively.
  + Streamlined Training: By implementing hyperparameter tuning and k-fold cross-validation, the proposed system streamlines the training process, facilitating faster convergence and swift deployment in real-world scenarios.
  + Addressing Class Imbalance: Effectively mitigating class imbalance issues within the dataset, the system employs oversampling, undersampling, or class weighting techniques, reducing bias towards the majority class and enhancing predictions for rare skin cancer types.
  + Clinical Impact: The system's enhanced accuracy and efficiency may positively impact patient care by enabling early and accurate skin cancer diagnoses, potentially leading to improved treatment outcomes and patient survival rates.

### SYSTEM DESIGN

* 1. System Architecture



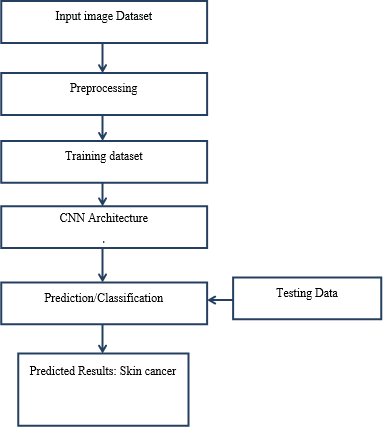
## Figure 7.1 System Architecture

* 1. Data Flow Diagram

The Data Flow Diagram (DFD), often referred to as a bubble chart, is a visual representation that simplifies the depiction of a system. It illustrates the flow of input data into the system, the processing carried out on this data, and the resulting output data. As one of the fundamental modeling tools, the DFD is essential for understanding system components. These components include system processes, the data manipulated by these processes, external entities interacting with the system, and the flow of information within the system.

Through graphical representation, the DFD elucidates how information traverses the system and undergoes transformations. It illustrates the movement of data from input sources to output destinations, highlighting the modifications applied along the way.

The DFD, synonymous with a bubble chart, offers a versatile framework for representing systems at various levels of abstraction. It can be segmented into different levels, each representing a higher degree of information flow and functional detail within the system.



## Figure 7.2 Data Flow Diagram

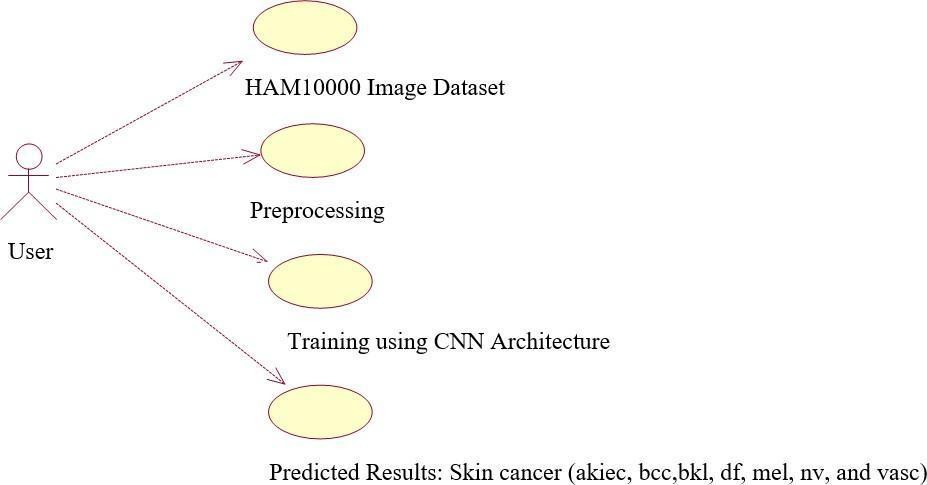
* 1. UML Diagram

UML, which stands for Unified Modeling Language, is a standardized modeling language widely used in the domain of object-oriented software engineering. Managed by the Object Management Group, UML aims to serve as a universal language for creating models of object-oriented software systems. Currently, UML comprises two main components: a meta-model and a notation.

The design of UML is driven by several primary goals:

1. Offer users a readily accessible, expressive visual modeling language to facilitate the development and exchange of meaningful models.
2. Provide mechanisms for extendibility and specialization to enhance core concepts.
3. Maintain independence from specific programming languages and development processes.
4. Establish a formal foundation for comprehending the modeling language.
5. Foster the growth of the object-oriented tools market.
6. Support advanced development concepts such as collaborations, frameworks, patterns, and components.
7. Integrate best practices to enhance modeling effectiveness.
   1. UseCase Diagram

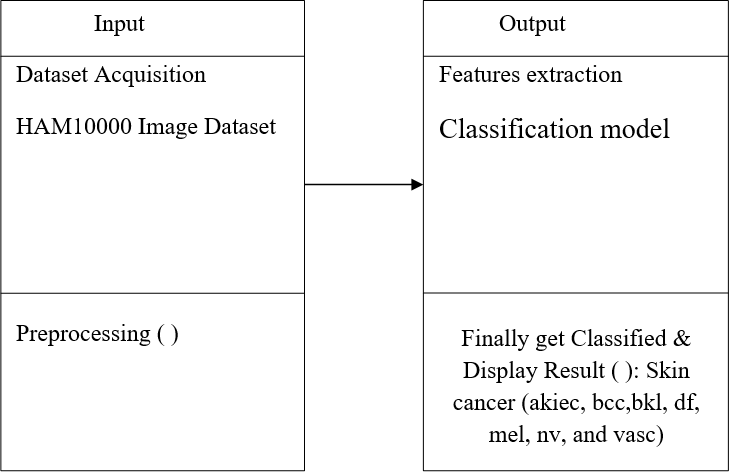
A Use Case Diagram, within the Unified Modeling Language (UML), is a behavioral diagram derived from Use-Case Analysis. Its primary function is to offer a visual representation of the system's functionality, delineating actors, their objectives (presented as use cases), and any interdependencies among these use cases. The principal aim of a use case diagram is to illustrate the system functions executed for each actor. Additionally, the diagram can portray the roles of actors within the system.



## Figure 7.4 Use Case Diagram

* 1. Class Diagram

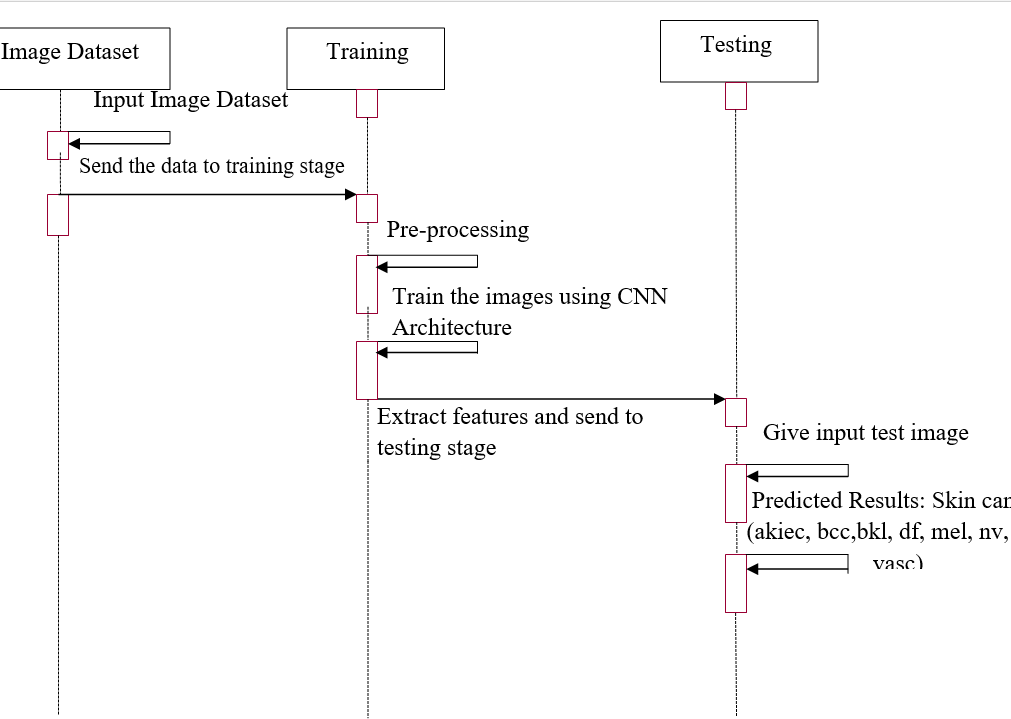
In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information.



## Figure 7.5 Class Diagram

* 1. Sequence Diagram

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams.

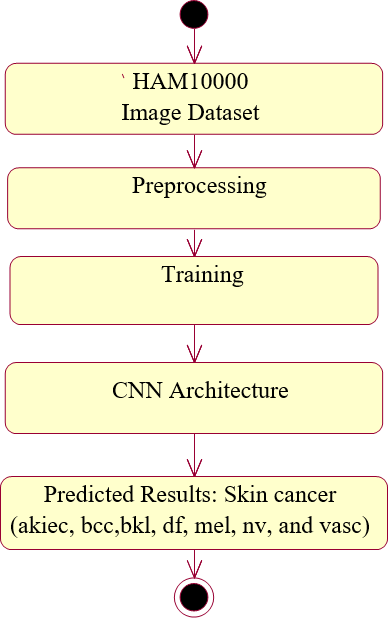


## Figure 7.6 Sequence Diagram

* 1. Activity Diagram

Activity diagrams are graphical representations of workflows of stepwise activities and actions with. support for choic iteration and concurrency. In the Unified Modeling Language, activity diagrams can

be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control



## Figure 7.7 Activity Diagram

### SYSTEM REQUIREMENTS

* HARDWARE REQUIREMENTS:
  + System: Pentium i3 Processor.
  + Hard Disk : 500 GB.
  + Monitor: 15’’ LED
  + Input Devices: Keyboard, Mouse
  + Ram:6 GB

### SOFTWARE REQUIREMENTS:

* Operating system: Windows 10 Pro.
* Coding Language: Python 3.10.9
* Web Framework: Flask
  1. RESULTS & DISCUSSION

## Table 9. Results and Observations Test Cases

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Test Case | | | Result/Observation | | Screenshot |
| Home | o | t | The screenshot | |  |
| Page | f | h | attached shows the | |
| Website |  | e | first interface the user | |
|  |  |  | shall experience when | |
|  |  |  | he opens the website. | |
| Scrollable: |  |  | The | ima |  |
| Home Page Website | o f | t h e | following displays  the | ge  low er |
|  |  |  | section | our |
|  |  |  | of |  |
|  |  |  | homepage. |  |
| Login Page | | | This is the Login Page for the User that asks for their details as shown. | |  |

|  |  |  |
| --- | --- | --- |
| Login Success | This is the successful login page that prompted the user's details. |  |
| Upload Image | This page prompts users to upload an image for skin cancer prediction. |  |
| Image Test | The Image Test Page is generated  upon uploading an image; with the image specifically selected/uploaded by the user. |  |

### CONCLUSION

In summary, the proposed system harnesses the capabilities of convolutional neural networks while integrating various enhancements to enhance accuracy, efficiency, and generalization capabilities. With a training accuracy of 96.00% and validation accuracy of 97.00%, our system demonstrates robust performance. By meticulously preprocessing data and organizing datasets, the project adeptly manages diverse dermatoscopic images and various skin cancer types. The incorporation of transfer learning and data augmentation techniques further fortifies the system against variations in image quality and lighting conditions, mitigating overfitting concerns.

The project's modular design ensures a structured approach, facilitating seamless integration and reusability. Each module plays a pivotal role in data handling, model construction, training, evaluation, and result analysis. Evaluation on a separate test set showcases the system's high accuracy in distinguishing between different skin cancer types. Moreover, interpretability techniques provide transparency into the model's decision-making process, fostering trust among healthcare professionals.

With real-time inference capabilities, the proposed system is poised for deployment in clinical settings, potentially expediting skin cancer diagnoses. Overall, this project represents a significant advancement in leveraging deep learning for skin cancer prediction. By amalgamating cutting-edge technologies and domain-specific knowledge, the system offers an efficient and reliable tool for early detection and classification of skin cancer types, potentially improving patient outcomes and alleviating the global burden of this disease.

While promising, further validation and benchmarking against larger and more diverse datasets are imperative to bolster the system's applicability and robustness. Continued research and development in medical image analysis and deep learning are poised to yield even more sophisticated predictive models in the future. Nevertheless, this project lays a solid foundation for future advancements in skin cancer prediction, benefiting medical professionals and patients alike in combating this pervasive and potentially fatal ailment.

### FUTURE SCOPE OF THE WORK

While the current project on skin cancer prediction, employing deep learning techniques, has made significant strides in accuracy and efficiency, there remain several avenues for future enhancement:

1. Expansion of Datasets: Enlarging the dataset to encompass a broader array of dermatoscopic images from diverse populations and skin types would enhance the model's generalization across various patient demographics.
2. Hyperparameter Optimization: Further exploration of hyperparameter tuning techniques, including automated optimization algorithms like Bayesian optimization or genetic algorithms, can optimize the deep learning model's configurations.
3. Explainable AI Techniques: Integrating advanced model interpretability techniques offers clearer insights into the features guiding the model's decisions. Techniques like SHAP (SHapley Additive exPlanations) or LRP (Layer-wise Relevance Propagation) provide a deeper understanding of the decision-making process.
4. Clinical Validation: Rigorous clinical validation studies on real-world patient data, conducted in collaboration with medical professionals, are vital to assess the system's performance in clinical settings. This ensures reliability and safety in practical healthcare applications.

### REFERENCE

1. A. Javaid, M. Sadiq, and F. Akram, “Skin cancer classification using image processing and machine learning,” in Proceedings of the 2021 International Bhurban Conference on Applied Sciences and Technologies,

pp. 439–444, IEEE, Islamabad, Pakistan, 12-16 Jan 2021.

1. P. Dubal, S. Bhatt, C. Joglekar and S. Patil, "Skin cancer detection and classification," 2017 6th International Conference on Electrical Engineering and Informatics, Langkawi, Malaysia, 2017, pp. 1-6, DOI: 10.1109/ICEEI.2017.8312419.
2. Emara, “A Modifier Inception\_v4 for imbalanced skin Cancer Classification”, 2019 14th International Conference on Computer Engineering and Systems, pp. 28-33, December 2019.
3. Y Jusman,” Performance of Multi-Layer Perceptron and Deep Neural Networks in Skin Cancer Classification”, IEEE Transactions, vol 10, pp. 118198 -118212, 2022.
4. R Raja Subramanian, “Skin Cancer Classification using CNN”, IEEE Conference, 2021 11th International Conference on Cloud Computing, Data Science & Engineering 10.1109/Confluence51648.2021.9377155, 2021.
5. Harikrishna, “Skin Cancer Classification using Transfer Learning IEEE International Conference on Advent Trends in Multidisciplinary Research and Innovation (ICATMRI-2020)
6. Nourabuared, “Skin Cancer Based on VGG19 and Transfer Learning”, In 3rd International Conference on Signal Processing and Information Security IEEE Conference, pp 1-4, DOI-https://doi.org/10.1109/ICSPIS51252.2020.93401432021.
7. Enakshi Jana, “Research on Skin cancer Cell Detection using Image Processing”, 2017 IEEE International Conference on Computational Intelligence and Computing Research, DOI: 10.1109/ICCIC.2017.8524554. pp. 1-8, 2017.
8. N. J. Dhinagar and M. Celenk, "Analysis of regularity in skin pigmentation and vascularity by an optimized feature space for early cancer classification," 2014 7th International Conference on Biomedical Engineering and Informatics, Dalian, China, 2014, pp. 709-713, DOI: 10.1109/BMEI.2014.7002865.
9. P. Sedigh, R. Sadeghian, and M. T. Masouleh, "Generating Synthetic Medical Images by Using GAN to Improve CNN Performance in Skin Cancer Classification," 2019 7th International Conference on Robotics and Mechatronics, Tehran, Iran, 2019, pp. 497-502, DOI: 10.1109/ICRoM48714.2019.9071823.
10. Dwaipayan Choudhury, “Texture and Color Feature Based WLS Framework Aided Skin Cancer Classification using MSVM and ELM”, 2015 Annual IEEE India Conference, December 2015.
11. A. Masood and A. Al-Jumaily, "Semi-advised learning and classification algorithm for partially labeled skin cancer data analysis," 2017 12th International Conference on Intelligent Systems and Knowledge Engineering, Nanjing, China, 2017, pp. 1-4, DOI: 10.1109/ISKE.2017.8258767.
12. N. C. Lynn and Z. M. Kyu, "Segmentation and Classification of Skin Cancer Melanoma from Skin Lesion Images," 2017 18th International Conference on Parallel and Distributed Computing, Applications and Technologies, Taipei, Taiwan, 2017, pp. 117-122, DOI: 10.1109/PDCAT.2017.00028.
13. Chee Jen Ngeh, “Deep Learning on Edge Device for Early Pre-Screening of Skin Cancers in Rural Communities, 2020 IEEE Global Humanitarian Technology

Conference, DOI:10.1109/GHTC46280.2020.9342911, October 2020.

1. S. R. Guha and S. M. Rafizul Haque, "Convolutional Neural Network Based Skin Lesion Analysis for Classifying Melanoma," 2019 International Conference on Sustainable Technologies for Industry 4.0 (STI), Dhaka, Bangladesh, 2019, pp. 1-5, doi: 10.1109/STI47673.2019.9067979.