# A Comparative Analysis of Deep Learning and Hybrid Models to Diagnose Multi-Class Skin Cancer

by

Ishrat Nur Nawrin
19301160

Tonusree Talukder Trina
19301158

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
Brac University
May 2023

# Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

**Student's Full Name & Signature:**

Ishrat Nur Nawrin

19301160

Tonusree Talukder Trina

19301158

# Approval

The thesis titled "A Comparative Analysis of Deep Learning and Hybrid Models to Diagnose Multi-Class Skin Cancer" submitted by

1. Ishrat Nur Nawrin (19301160)
2. Tonusree Talukder Trina (19301158)

Of Spring, 2023 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on May 25, 2023.

**Examining Committee:**

Supervisor:
(Member)

Annajiat Alim Rasel

Senior Lecturer
Computer Science and Engineering
Brac University

Co-Supervisor:
(Member)

Rafeed Rahman

Lecturer
Computer Science and Engineering
Brac University

Program Coordinator
(Member)

Md. Golam Rabiul Alam

Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Ms. Sadia Hamid Kazi

Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

# Ethics Statement (Optional)

This is optional, if you don't have an ethics statement then omit this page

# Abstract

Skin cancer is one of the most lethal and increasingly prevalent cancers in the world. Skin cancer develops when the epidermal (top layer of skin) cells divide abnormally, causing it to spread to other regions of the human body. Skin cancer exists in seven different varieties. The presence of malignant epidermal cells determines the type of skin cancer. Dermoscopy, spectroscopy, and imaging tests are primarily utilized to identify the malignancy. These procedures are expensive and prolonged. It may result in unfavorable effects such as bleeding, bruising, and infection as well. The narrow variances in multi class cancer pictures escalate the complexity of classification. Dermatologists confront challenges in the categorization of cancer types from images. Deep learning has resulted in a dramatic leap in disease identification. Deep learning models are capable of categorizing skin cancer more precisely than dermatologists. Several studies focused on pre-trained and hybrid models for categorizing the classes of skin cancer. In contrast to binary classification, the multi-class classification of skin cancer yielded an insignificant result for both deep learning and dermatologists. The proposed study employs varieties of deep learning and hybrid models to examine the performance of each model in categorizing the classes of cancer. The proposed CNN-SVM-LSTM hybrid model obtained the highest result compared to other models, with 87.15% accuracy, 87.42% precision, 87% recall, and 87.428% F1 score. To illustrate the overall comparison of the models, each model has been depicted through a classification report and a confusion matrix.

**Keywords:** hybrid models; lethal; classification; prolonged

# Acknowledgement

Firstly, all praise to the Great Allah for whom our thesis have been completed without any major interruption.
Secondly, We are thankful to our supervisor, Annajiat Alim Rasel sir, and co-advisor, Rafeed Rahman sir, for their kind support and advice in our work. They always assisted us with appropriate instruction.

# Table of Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 Motivation

In the preceding timeframe, the dominance of machine learning and deep learning in medical disease detection and classification has emerged in a short span of time. Deep learning has surpassed humans in disease detection in terms of precision and accuracy. Failing to detect skin cancer early accelerates the risk and shrinks the chance of survival. Even though machine learning and deep learning demonstrated outstanding performance in the binary classification of skin cancer, several problems arise while detecting classes of cancer due to the fine differences between the classes. An experienced dermatologist also faces a challenge when classifying the type of cancer. The procedure of treatment cannot be ensured due to the failure to detect the type of cancer. The proposed study focuses on comparing several models along with a few customized models to find the most suitable model for classifying cancer.

## 1.2 Aim and Objective

The purpose of the study is to analyze the pre-trained and customized hybrid models to illustrate a comparison of their performance. The models were mapped based on their performance in detecting skin cancer types. The proposed hybrid models demonstrated remarkable performance compared to the traditional models. The comparison has been illustrated using some parameters like the F1 score, precision, and recall.

## 1.3 Contribution

To reach the goal of our research, several strategies were adopted. A few new strategies contributed to attaining the goal, which is noted down below to comprehend the contribution of the research.

- The study introduced the idea of using k-fold cross-validation to provide a robust approximation regarding the performance of the model. K-fold validation creates k segments of data, and each segment is processed using the model. The repeated evaluation of models narrows the variation in performance and ensures the integrity of the models.

- The proposed six-hybrid model was evaluated to identify the leading performer model of the comparative study. The proposed models are a fusion of convolutional neural networks and deep learning models. The proposed fusion models contributed to accomplishing the goal of the research.

- We opted for the idea of inducing bidirectional LSM layers to construct a hybrid model. Bidirectional LSTM is a vital component in sequential recognition domains. A bidirectional LSTM was used in the study to preserve contextual information. The model produced an exceptional outcome.

## 1.4 Problem Statement

The skin is the protective organ in the human body that shields the body from a variety of infections and diseases. This defensive organ is also susceptible to infection. Human skin can be categorized into three layers: the epidermis on top, the dermis in the middle, and the hypodermis at the bottom. Cancer develops in the epidermis, which serves as the body's protective barrier. Skin cancer is classified mainly into two categories: non-melanoma and melanoma. Melanoma is the most life-threatening skin cancer type, with a higher mortality rate. Basal cell carcinoma and squamous cell carcinoma are the two most frequent kinds of non-melanoma skin cancer. Although basal cell carcinoma and squamous cell carcinoma can be fatal, they have a higher survival rate than melanoma skin cancer.

The epidermis layer constantly produces new skin cells. The newly produced cells can replace nearly 40,000 old skin cells every 30 days [16]. When this procedure fails, a rapid increase in cells occurs, resulting in skin cancer. Ultraviolet (UV) rays are the leading cause of skin cancer. It damages the DNA of skin cells, which leads to a state called aberrant cell proliferation. This condition is a notable feature of skin cancer. Cyclobutane Pyrimidine Dimers and Photoproducts are two forms of UV-induced DNA damage. These damages cause distortions in DNA's structural integrity. In the beginning, UV light induces DNA damage in epidermal keratinocytes, which leads to the mutation process. This mutation results in a condition known as "carcinogenesis" in skin cells. In this stage, tissue cells replicate uncontroversially, causing normal cells to become cancerous [3]. Melanin is a skin pigment that absorbs harmful UV radiation and minimizes cellular damage. People with low melanin are at the highest risk of developing skin cancer. Skin with low melanin is incapable of absorbing harmful UV rays and fails to shield the body from cellular damage. A study published in the JDNA (Journal of the Dermatology Nurses' Association) [13] reports that having five or more sunburns between the ages of 15-20 raises the risk of melanoma skin cancer by 80%. Sunburn during childhood or adolescence enhances the risk of developing melanoma skin cancer by twofold in light-skinned people.

Australia, New Zealand, the United States, and European countries are the most affected by skin cancer [8]. Lack of the protective pigment melanin in light-skinned people raises their cancer risk. Skin cancer can strike people with darker skin as well. Darker-skinned people have more epidermal melanin and melanocyte activity, which allows them to filter UVB radiation twice. Despite having melanin, they are still at high risk of skin cancer due to late detection. This is because their skin cancer symptoms are more difficult to detect. A late diagnosis of skin cancer raises

the risk and can be fatal. The high medical expense of cancer detection is another factor in second- and third-world nations' failure to detect and diagnose skin cancer at the primary stage. Hence, it is critical to detect skin cancer at the earliest possible stage.

# Chapter 2

# Literature Review

## 2.1 Background Study

This segment highlights the concepts, theories, and background knowledge of research to comprehend the relevance of the proposed models. This section covers CNN, SVM, LSTM, hybrid and pre-trained models and their contributions to the proposed hybrid model implementation based on perceived knowledge.

### 2.1.1 Convolutional Neural Network (CNN)

Convolution Neural Network (CNN) is the most extensively used model for computer vision and medical disease diagnosis. It has demonstrated promising results across a wide range of fine-grained item categorizations in a variety of broad and highly variable tasks [17]. CNN utilizes filters to extract features and learn complex features along with the depth of the convolution layer. Based on the extracted features, the model learns and enables it to categorize the unknown images. CNN has the potential to provide impressive results without redundant computation .The performance of the CNN model surpasses other deep learning models in the domain of image classification [6]. The output equation in CNN architecture is denoted as , The sign X refers to the input of the network. The term W refers to the weight of

$$Y = F(\sum_{i=1}^{N} = (X * W) + b$$

the corresponding input, and b indicates the bias of the input.

### 2.1.2 Long Short Term Memory (LSTM) Network

LSTM is a variation of the Recurrent Neural Network (RNN) that recognizes sequential structure by following long-term dependencies. The characteristic of long-range dependencies in features makes LSTM applicable for image classification as well. In some scenarios, LSTM hybrid models perform better than deep learning models [16]. The feedback structure in LSTM enables it to identify a sequential pattern from a single point of data. LSTM is an updated version of RNN that overcomes the vanishing gradient and exploding problems of RNN.

**Overview of LSTM Cell**

In LSTM architecture, each LSTM cell comprises three gates, which are the input gate, forget gate, and output gate. The gating mechanism of LSTM makes it capable of controlling the flow of data. The given equations denote the processed outcome provided by the gates of LSTM. The expression $h_t - 1$ denotes the output of the previous state, and $x_t$ is the current input value. These two equations are

$$i_t = \sigma(W_i.[h_t] - 1, x_t) + b_i....(1)$$

$$C_t = tanh(W_c.[h_t] - 1, x_t) + b_c....(2)$$

accomplished by the outcome of the input gate. The term $i_t$ indicates the input of the cell; the term w symbolizes the weights of the corresponding input; and b refers to the bias of the input value. Equation 3 represents the outcome of the forget gate.

$$f_t = \sigma(W_f.[h_t] - 1, x_t) + b_f....(3)$$

The gate uses the sigmoid function as activation and provides a value in the range of 0 to 1.

The output of the LSTM cell is demonstrated through equation 4, where $h_t$ is the

$$O_t = \sigma(W_o.[h_t] - 1, x_t) + b_o....(4)$$

$$h_t = o_t * tanh(C_t)$$

next state value and $O_t$ is the received output.

### 2.1.3 Support Vector Machine

Among the variety of classifiers, SVM is one of the most remarkable due to its key functionality of utilizing linear functions in high-dimensional feature space. SVM follows a supervised approach that is extensively used both in regression and classification tasks [1]. The features of a data set can be differentiated into linear or non-linear structures. The linear structure pattern can be separated by using a single line. Nonlinear features cannot be distinguished with a single line. The data points need to be manipulated to separate the classes [2]. The classifier implies the concept of finding a hyperplane to differentiate the classes. The classifier maps the n-dimensional input into high-dimensional feature spaces [4]. In the scenario of multi-class data, SVM adopts the one-versus-all approach or the one-versus-rest approach. Images are basically a representation of a high-dimensional vector, which can be classified smoothly by the plane of the SVM classifier. In the field of pattern recognition, SVM achieved better reliability and proved to be the most effective technique compared to other models [2].

### 2.1.4 Random Forest Classifier

Random Forest is an ensemble model that maintains a structure similar to a tree. The classifier predicts the outcome based on the concept of voting using numerous

decision trees. It uses multiple decision trees and the output is finalized depending on the majority voting from decision trees **chaudhary2016**. The classifier has the ability to identify non-linear structures in the data. The working principle of the random forest classifier is illustrated below in the image.

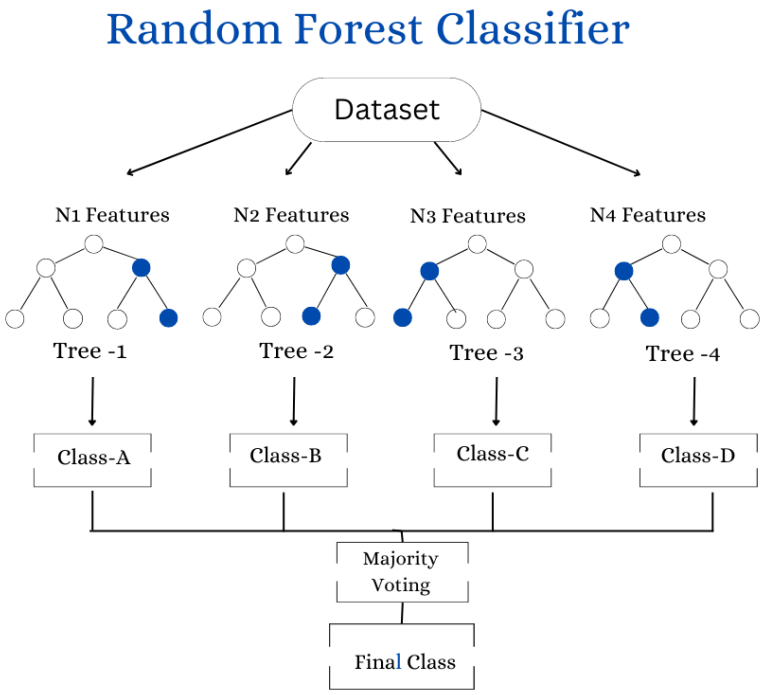


Figure 2.1: Random Forest Classifier

### 2.1.5 Extreme Gradient Boosting (XGBoost )

XGBoost is a form of gradient boosting that is faster and more systematic than gradient boosting methods. By maintaining a sequential structure, it generates a quicker result. In the XGBoost tree structure, the trees are structured in a sequential order, and the child trees keep correcting the errors of the previous tree. The principal advantage of the XGBoost algorithm is that it can handle missing values. It can process the values in parallel as well. The parallel processing structure makes the algorithm faster even for large datasets [22]. Another notable feature of the XGBoost classifier is that it is scalable in all scenarios, which makes it suitable for real-world applications. The XGBoost algorithm runs 10 times faster compared to other algorithms and scales to billions of distributions [5].

### 2.1.6 Pre Trained Models

#### DenseNet

DenseNet is a popular transfer learning model for medical image classification. It consists of 120 convolution layers and four average pooling layers, all with the same dense block. The dense connection in the network allows it to maintain direct connection between the blocks. In this network, the output of the preceding cell is

carried out as the input of the next layer.. The feature map created at each layer is transmitted to the next layer as an input [7]. This system diminishes the vanishing gradient problem which occurs as a result of the distance between the input and the output layer. It provides better performance than other models due to its functionality to solve vanishing gradient problems. DenseNet 121 has outperformed all other transfer learning methods for detecting skin cancer.

**AlexNet**

AlexNet is the first CNN based winner in the competition of ImageNet Large-Scale Visual Recognition Challenge (ILSVRC). The architecture of AlexNet encompasses 5 convolutional neural networks which performs the role of extracting the hierarchical features from the input images. The convolution layers are adjacent to the Relu layer. The Relu layers are capable of learning complex features from the input. To downsample the images network employs 3 max pooling layers . In this model 3 fully connected layers have been incorporated and this layer utilizes more parameters compared to the Conv layers. To diminish the over-fitting of data a regularization approach called dropout has been employed [9].

**VGG 19**

VGG19 is a pre-trained convolutional neural network. It encompasses 16 convolutional layers, with three fully connected layers. The model preserves the spatial resolution after each convolution operation and downsamples the images only at the max pooling stage. VGG19 is an upgraded version of VGG16 that also overcomes the limitations of the AlexNet architecture. It demonstrated a more accurate result in the field of image classification than other deep learning networks like AlexNet, VGG16, and other hybrid models [9].

**ResNet**

Resnet50 is an improved form of the ResNet (Residual Network) Model. The network is composed of 50 layers, which include convolutional layers, pooling layers, and fully connected layers. It has the potential to train extremely deep or shallow networks. The architecture has a similar structure to VGG16, with residual blocks. The residual connection in the network introduced skip connections, which allow the model to use feature maps in numerous positions between the layers [21]. The concept of creating skip connections helped the ResNet model learn and maintain its integrity.

## 2.2 Related Work

Many computerized-based cancer detection and recognition algorithms have been employed in the past to determine illnesses and anomalies in skin lesions.
A study highlights a comparison between dermatologists and convolutional neural networks in terms of detecting the class of cancer [10]. The systematic comparative study reveals that convolutional neural networks outperform dermatologists. The systematic study emphasizes two main objectives. The primary objective of the study is to distinguish between the binary classifications of cancer (malignancy or

non-malignancy). Another objective is to categorize the classes of cancer into five different categories. A group of 112 dermatologists from 13 different well-known German universities participated in the comparative study. For testing purposes, 300 images were analyzed by dermatologists. The samples were also used in CNN models to diagnose the cancer type. After examining the performance, it was noticeable that CNN outperformed the dermatologist in both objectives of the comparative study. The study also demonstrated that both CNN models and dermatologists tend to misclassify melanoma as nevi. The overall comparison reveals that the performance of CNN is more advanced than the performance of the experienced dermatologists.
A study proposed to classify malignant and nonmalignant types of melanoma cancer by following three phases [11]. The research was conducted on the ISIC dataset. In the initial phase, the data went through pre-processing steps, which included hair removal (using Hough transform), shading removal, and glare removal. The pre-processed images were taken to the second phase for feature extraction and segmentation. Three forms of segmentation were inspected for the research: the Otsu segmentation method, the modified Otsu segmentation method, and the Watershed segmentation method. For feature extraction, color, shape, and size were retrieved as features. In the final step, the processed data were fed to three model-back propagation algorithms: neural networks, support vector machines (SVM), and convolutional neural networks (CNN). The SVM model demonstrated incredible performance compared to the other two models.
In the binary classification of cancer, the convolution network and deep learning networks achieved remarkable outcomes. Multi-class skin cancer classification is comparatively challenging and complex due to the subtle distinctions between classes. Regardless of the complexity, several deep learning models demonstrate significant performance in multiclass cancer diagnosis. A proposed study [20] used modified EfficientNet B0-B7 models to classify cancer types and analyze the performance of different models. A fine-tuning strategy was used to improve the outcome. It was observed that high complexity does not always perform best. The intermediate complex model EfficientNet B4 outperformed the other EfficientNet complicated models with an accuracy of 87.95%.
Another study emphasized constructing a benchmark for skin cancer that will be able to assess the OOD (out of data) of the classifiers [18]. This paper utilizes multiple datasets, including the HAM10000, ISIC, BCN20000, PH2, DERM 7PT, and SKINL2. The dataset was processed for training and validation without data augmentation. For testing purposes, the SAM, SAM-C and SAM-P datasets have been employed. The initial dataset SAM, comprises the original photos. The other two datasets, SAM-C and SAM-P, contain the SAM images which are intentionally distorted or processed using 22 different sorts of modification processes. To analyze the performance, four CNN architectures have been incorporated. The four CNN models are AlexNet, VGG16+BN, ResNet50, and DenseNet121. Among these 4 models, VGG16+BN and ResNet50 demonstrated the leading performance for SAM, whereas DenseNet121 is the supreme performer for SAM-C. For the dataset, SAM-P and AlexNet showed the highest performance.
A research proposal by Hameed, Shabut, and other authors [14] presents the implementation of a "Multi-Class Multi-Level (MCML)" classification algorithm to categorize skin lesions. The research is conducted on a combination of ISIC and PH2 datasets. The examination was administered in two stages. In the first stage,

the researcher made use of a machine learning algorithm, whereas in the later stage, a deep learning algorithm was employed. The result from the experiments was compared with the "multi-class single-level" classification algorithm. By examining the outcome, the authors claim that the MCML classification algorithm contributes to escalating the performance of the classification of multi-class skin lesions.
In 2021, Aggarwal et al.[19] aimed to diagnose melanoma and basal cell carcinoma on the skin of diverse groups. The study aims to construct AI models that could detect cancer in different skin tones. The researcher collected images of basal cell carcinoma and melanoma types from dermatological databases. As the research emphasized diverse skin tones, they darkened the images through a method called Fast Contrastive Unpaired Translation (FastCUT). They performed the process on 553 basal cell carcinoma images, 324 melanoma images, 261 BCC images, and 102 melanoma images. To carry out the research, both light-skinned and dark-skinned images were trained on two deep learning models: CNN and Inception-ResNet-V2. The researcher finds that for the darker-skinned images, the CNN model performs better than for the images with a light skin tone.
Another research study utilized modified deep learning models and ensemble models to improve performance [12]. They experimented on pre-trained models including Xception, InceptionV3, InceptionResNetV2, ResNeXt101, and NASNetLarge. To improve the existing model, they included Rectified Linear Unit (ReLU) in the dense layer, placed the dropout layer and softmax layer at the bottom of the architecture, and tweaked the parameter values. ResNeXt101 outperformed all deep learning models with an accuracy of 93.2%. The ensemble model, on the other hand, obtained 92.83% accuracy.
A paper published in 2021 [15], demonstrated the classification of melanoma with the help of a recognition system in two systematic approaches. The researcher collected melanoma images from the PH2 and ISIC 2018 datasets. The collected images were processed using gaussian filters to remove noise and blur the images. In the first developed system, the researchers extracted the features from the processed images by applying the Local Binary Pattern (LBP) and Gray Level Co-occurrence Matrix (GLCM). The retrieved features were tested on an ANN model. In the other approach, the images were passed through a CNN model that had been previously pre-trained using AlexNet and Resnet transfer learning models. The comparison shows that the ANN model or the first approach surpasses the other model with remarkable accuracy.

## 2.3 Proposed Models

The comparative study is performed on a convolutional model, three popular pre-trained models, and six customized hybrid models. . The proposed fusion models are built on the concepts of deep learning and convolutional network collaboration.

### 2.3.1 CNN-SVM

Convolutional neural networks (CNN) and deep learning networks are the most prominent tools in the area of image classification. The SVM is another remarkable

tool for classification. In order to implement the CNN-SVM fusion model, the proposed CNN model has been reprocessed using k-fold cross validation. The acquired features from the CNN model were conveyed to the SVM classifier. The classifier performs the task of categorizing the data.

### 2.3.2 CNN-Xgboost

The combined model of CNN and Xgboost has the same structure as CNN-SVM. In this architecture, the Xgboost classifier was incorporated in place of the SVM classifier. The ensemble classifier was employed due to its functionality in boosting the performance of the model.

### 2.3.3 CNN-Random Forest Classifier

The CNN-Random Forest hybrid model composed of a CNN architecture and random classifier. The input images are extracted by the CNN model. K-fold validation ensures better learning of the model. The collected features are passed into the random forest classifier.

### 2.3.4 CNN-LSTM

The proposed CNN-LSTM fusion model was built using five convolutional layers, each adjacent to a batch normalization layer and a maxpooling layer. Following the last pooling layer, three LSTM layers with a dropout value of 0.05 and a recurrent dropout value of 0.20 were used. After modification of the image size to fit into the model, three dense layers with Relu activation were put next to the model. K-fold validation was performed for adequate learning. Four folds were used due to the limitations of the available memory in the PC. The default learning rate is synchronized at 0.001, and Adam is utilized as the optimizer.

### 2.3.5 CNN-LSTM-SVM

The proposed CNN-LSTM-SVM surpassed all the tested models with an accuracy of 87.15%. A similar structure as the CNN-LSTM model was followed to build this hybrid model. In this model, bidirectional LSTM layers were used instead of LSTM layers. Bidirectional LSTM is an extension of LSTM that allows input to flow in two directions. It preserves information from the forward flow and the backward flow of the model, making it more efficient than LSTM. After extracting features, the data gets flattened through the FC layer. The output from the FC layer generates an one dimensional array, which is then fed via the support vector machine layer.

# Chapter 3

# Methodology

## 3.1 Data Description

The multi class skin cancer dataset was obtained from the Kaggle. Ham10000 is a standard dermatological image dataset consisting of 10,015 samples of skin lesions. The dataset includes seven different categories of skin cancer, such as actinic keratosis (akiec), benign keratosis (bkl), dermatofibroma (df), melanoma (mel), melanocytic nevi (nv), basal cell carcinoma (bcc), and vascular lesions (vasc). This standard dataset is frequently used in machine learning networks for cancer detection. A significant fraction of the data belongs to the nevi class, which is responsible for making the dataset highly imbalanced. In contrast, the df class includes the least amount of data in the dataset. In order to dispose of this imbalance, two procedures have been utilized in this research: data augmentation and SMOTE.

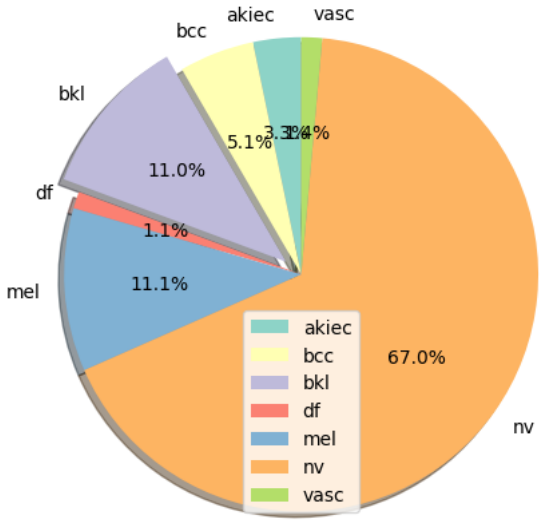


Figure 3.1: Ham10000 Dataset

## 3.2 Image Pre-processing Techniques

Image processing is the initial step to prepare the input for testing the performance of the model. The Ham10000 dataset is highly skewed and large. To obtain an

adequate result, minimal pre-processing strategies were adopted. The preprocessing method can be divided into three phases.

### 3.2.1 Data Augmentation

In the initial phase, a deep learning approach called data augmentation was adopted to increase the data of the smaller classes. This method creates new, slightly modified copies of the original data. Using this strategy, more unique and varied instances of the same data were provided in an effort to enhance model performance. The parameters of the ImageDataGenerator's rotation range, width shift, height shift, shearing images, zoom range, horizontal flipping of images, and fill mode were applied in order to balance the unbalanced data. The Figure 3.1 shows that the melanocytic Nevi class accounted for 67% of the data. The motivation behind performing data augmentation was to balance the skewed data. After applying data augmentation, a remarkable change was visible on the dataset. The Figure 3.2 reflects this outstanding modification. An illustration of data augmentation is shown in figure 3.3. As nv class already had reasonable data, data augmentation was not employed on this class.

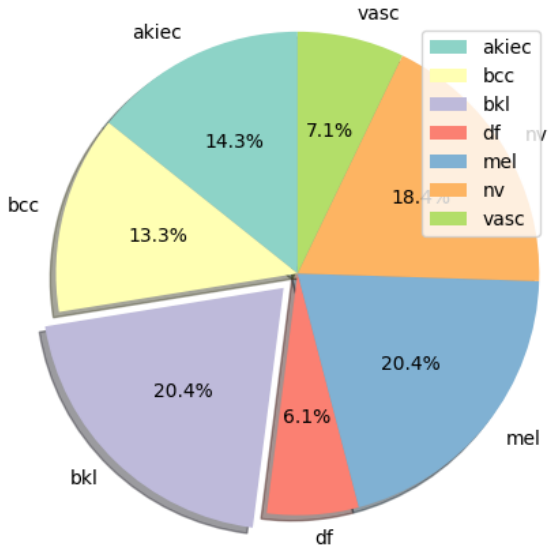


Figure 3.2: Dataset Distribution After Data Augmentation

### 3.2.2 Standardization

After performing data augmentation, the standardization method was utilized to scale the data. A standardization strategy is capable of efficiently storing information about outliers. This strategy reduces the algorithm's sensitivity. The study utilized the standardization strategy as the acquired data classes had significant differences in terms of the range of features. It is represented as,

$$X_{std} = \frac{x_i - \mu_x}{\sigma_x}$$

In this instance, the symbol mu expresses the distribution's mean. The symbol sigma symbolizes the standard deviation.

### 3.2.3 Synthetic Minority Oversampling Technique (SMOTE):

At the final step, the SMOTE method is applied to maintain a balanced range of classes. SMOTE is a statistical approach that generates synthetic points to balance the imbalanced data. Even after implying data augmentation, the dataset was not completely uniform. SMOTE was used to equalize the size of the data across all classes to balance the image classes.
After following the mentioned phases of preprocessing, the disparity in the range of data was balanced.

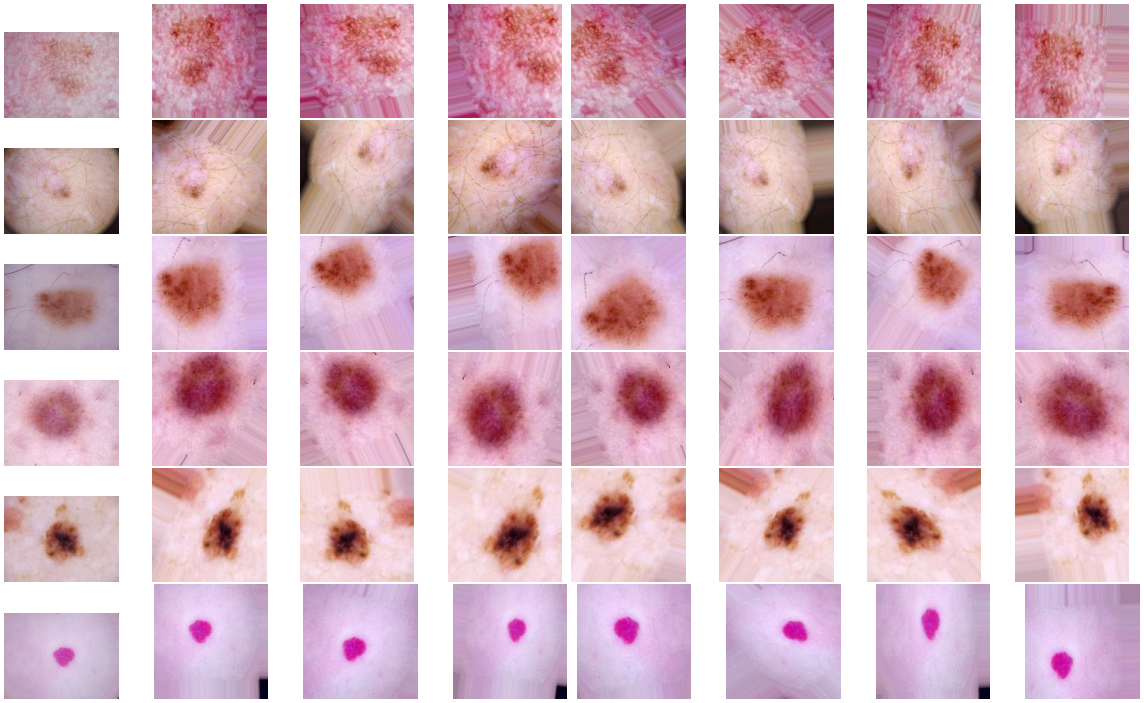

Figure 3.3: Data Augmentation

## 3.3 Model Assessment

**Pre-trained Model Assessment:**

Four transfer learning models, DenseNet121, VGG19, ResNet, and AlexNet, were employed from the Sci-kit library. The image size and batch size for each model vary according to the requirements of the models. The images were kept constant for the DenseNe121 and VGG19 models. On the other hand, the image size of 64 by 64 cannot be processed in ResNet or AlexNet. The size of the images varied according to their relevance to the model. To meet the image size requirement of AlexNet, it was set to 200 by 200 pixels. For the same purpose, in the ReNet50 model, the size is preserved as 150 by 150 pixels. Similarly, the batch size for ResNet50 and DenseNet121 was maintained at 32. In contrast, the batch size for AlexNet was 256 and 64 for VGG19.
**The average accuracy of the pre-trained models is in the range of 76-79%**

**Proposed model assessment:**

The research proposed six self-constructed models that exceeded the outcomes of the pre-trained models. The image shape of the input was kept constant at 64 by 64 pixels for each model. In the hybrid models CNN-SVM, CNN-XGboost, and CNN-Random Forest, the extracted features from the convolutional models were classified using the classifiers. On the other hand, the top-performing model CNN-Lstm-SVM is composed of convolutional layers, bidirectional LSTM layers, and an SVM classifier. The model took the most time to train compared to other models. The performance of each model is documented and compared with the other models.

**The mean performance of the proposed models was above 80%.**

**K-fold cross-validation:**

To ensure adequate learning of the data, the K-fold validation strategy has been applied. Due to the space limitation on the PC, the fold number was set to 4. Each fold of a model was trained, and an improved performance in validation was noticed for every fold.

## 3.4 Model Explanation

To train the dataset, a variety of deep learning models had been implied in this research including 1 customized, 5 customized hybrid, and 4 transfer learning models. After implying 6 customized models along with 4 transfer learning models, CNN-LSTM-SVM could outperform all the other models. This model could attain 87.15% accuracy. This hybrid model utilized a customized CNN and LSTM, with the final layer utilizing a Support Vector Machine.

1. **Customized CNN:** With the padding value continuing "same," five different layers had been employed for the customized CNN model, with the Relu activation function. Batch Normalization and Max Pooling were bundled to each layer.

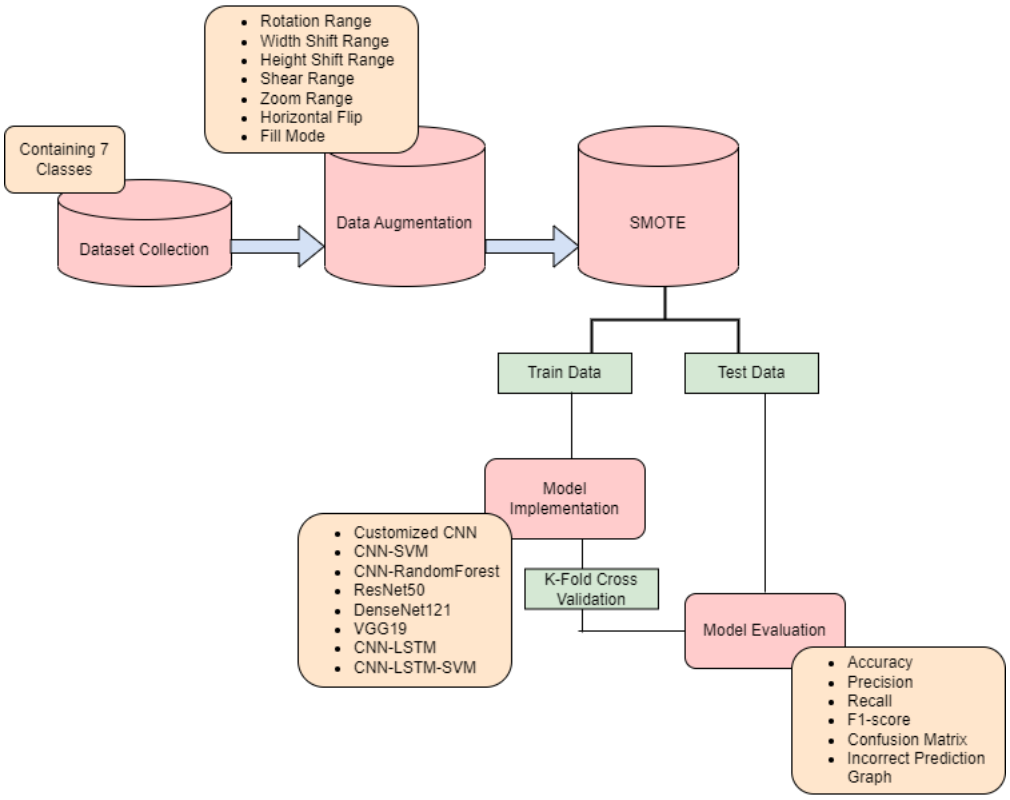


Figure 3.4: Workflow Diagram

2. **Customized LSTM:** Data was initially reshaped to fit in the LSTM layer for the customized LSTM model. With a dropout value of 0.05 and a recurrent dropout value of 0.20, 3 bidirectional LSTM layers had been added. In addition, 3 dense layers were added after 3 bidirectional LSTM layers using Relu as activation function. Data had been flattened to create the 1D array after these layers.

3. **Support Vector Machine:** Support Vector Machine was added as the last layer keeping the kernel regularizer value 0.01 and assigning activation function as Softmax. Regularization is the process of introducing penalty factors to the network layers to overhaul weights propagation through the layers, enabling the model to converge as accurately as possible.

K-fold cross validation was applied for the training of the customized hybrid model using the HAM10000 dataset. The batch size in this case was 256, and there were 4 folds. Here, default learning rate had been used which is 0.001 and Adam was used as optimizer.

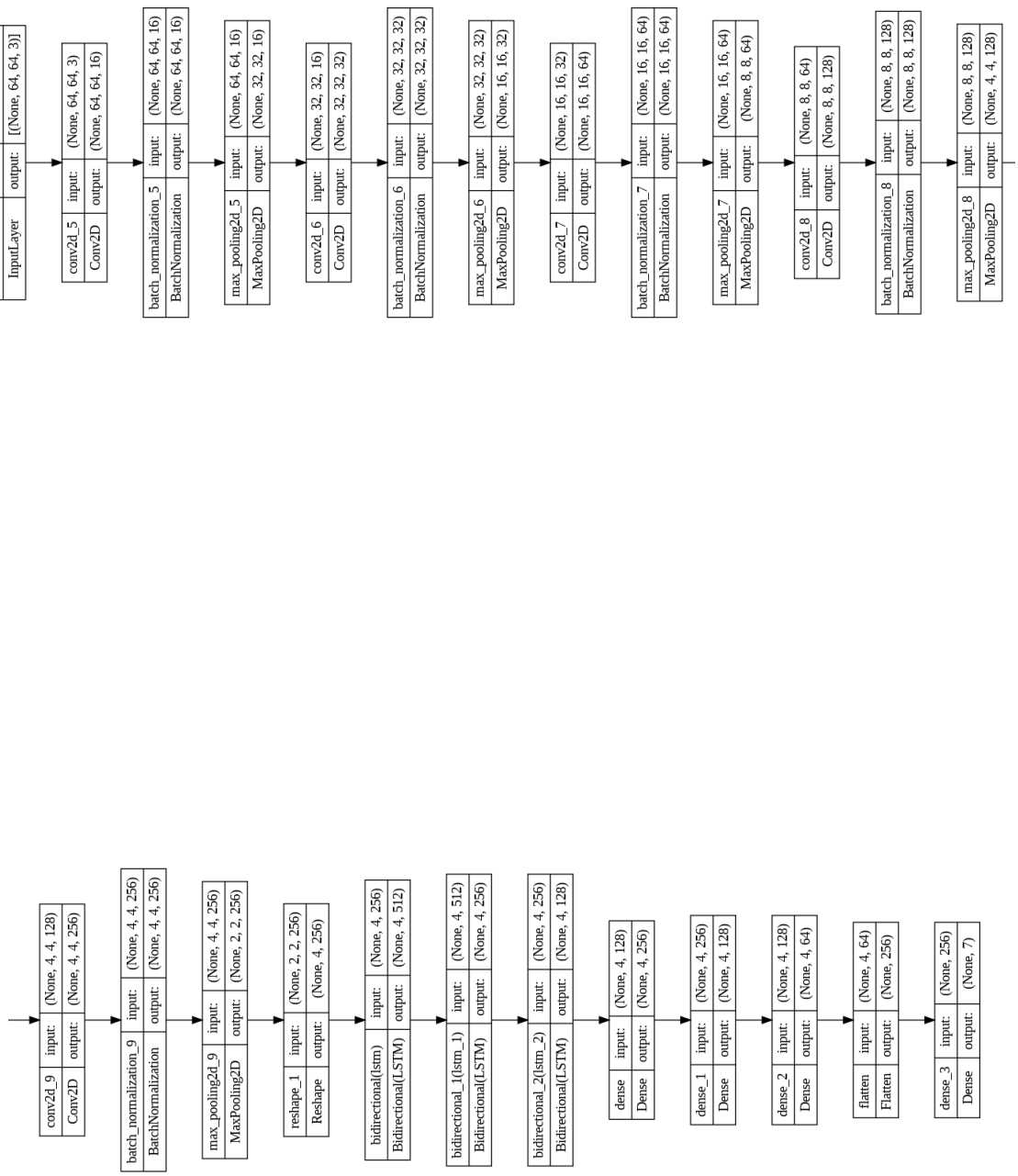


Figure 3.5: Summary of the Whole Architecture

# Chapter 4

# Result & Comparison

## 4.1 Evaluation parameters

The performance of a model can be depicted using parameters like F1 score, precision, recall, etc. The accuracy score, precision, recall, F1-score, and confusion matrix had all been plotted for each model to show how well it performed overall.
**Accuracy Score:** In classification problems, accuracy refers to the model's percentage of right predictions out of all predictions. It is described as the proportion of accurate predictions to all other predictions. **Precision:** The precision of a class

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

refers to the division of the total sample count by the true positive result proportion. True positive values are samples that have been assigned to a specific class. It forecasts the class using the computed frequency that was employed in the model.

$$Precision = \frac{TP}{TP + FP}$$

**Recall:** The term recall refers to the calculation of the true positive value divided by the sum of the true positive and false negative values. It evaluates the proportion of samples from that class that the model correctly recognizes.

$$Recall = \frac{TP}{TP + FN}$$

**F1-score:** A model's general efficacy is assessed using the F1-score, also referred to as the harmonic mean of recall and precision.

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

**Confusion Matrix:** The overall effectiveness of a model for a set of test data is displayed in a confusion matrix, a graphic representation table. There are four types of values: true positive (TP), true negative (TN), false positive (FP), false negative (FN). By computing measures like accuracy, precision, recall, and F1-score, these elements are utilized to assess the performance of an algorithm.

| Models | Image | Batch | Accuracy | Precision | Recall | F1 |
|---|---|---|---|---|---|---|
| Customized CNN | 64*64 | 16 | 85% | 85% | 85% | 85% |
| CNN-SVM | 64*64 | 16 | 86% | 86% | 86% | 86% |
| CNN-RandomForest | 64*64 | 16 | 84% | 84% | 84% | 84% |
| ResNet50 | 150*150 | 32 | 75.82% | 76% | 76% | 76% |
| DenseNet121 | 64*64 | 32 | 79.1% | 79% | 79% | 79% |
| VGG19 | 64*64 | 64 | 74.32% | 74% | 74% | 74% |
| AlexNet | 200*200 | 256 | 74.95% | 75% | 75% | 75% |
| CNN-XGBoost | 64*64 | 32 | 84% | 84% | 84% | 84% |
| CNN-LSTM | 64*64 | 32 | 82.78% | 83% | 83% | 83% |
| CNN-LSTM-SVM | 64*64 | 256 | 87.15% | 87% | 87% | 87% |

Table 4.1: A Summary of Evaluation of the Implemented Models

## 4.2 Performance table

Table 4.1 interprets the performance of each model along with the evaluation parameters. From the table, it appears that the performance of the proposed models surpassed the performance of the conventionally trained models. Except for ResNet50 and ALexNet, the image size was kept constant at 64 by 64 pixels. Compared to other pre-trained models, DenseNet achieved higher accuracy.

## 4.3 Comparison Analysis

The evaluated parameters of the models present a transparent view of the variety in the models performance. A systematic map (Figure 4.1) is created based on the models performance to perceive the hierarchical order of the models based on their performance. From all the implemented models, the proposed CNN-LSTM-SVM obtained the highest accuracy. The second top-performing model is the proposed CNN-SVM hybrid model. The VGG-19 demonstrated the least accuracy due to its shallow architecture.

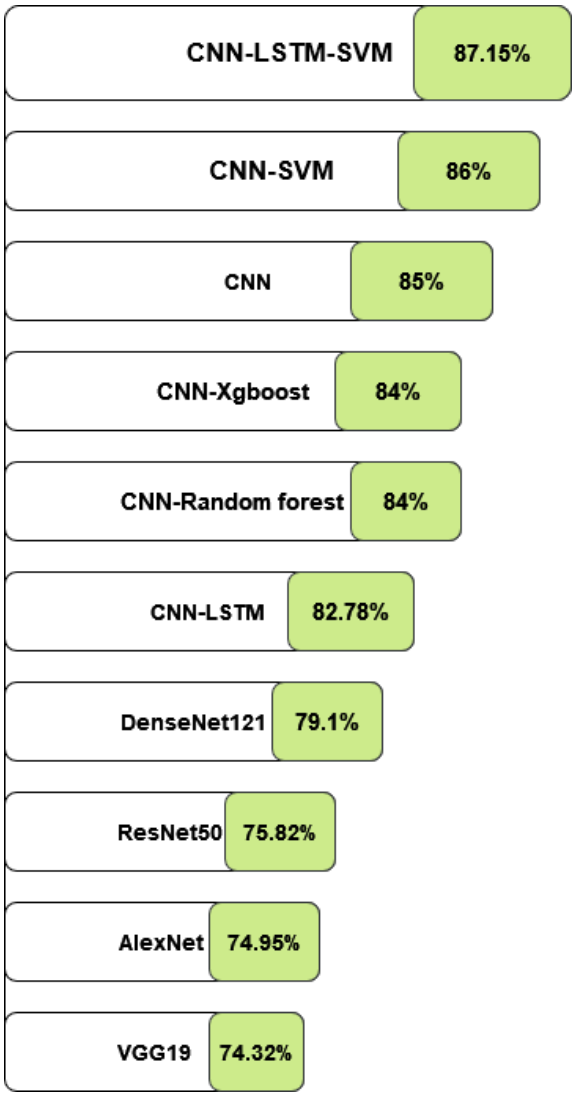


Figure 4.1: performance-based hierarchical structure

### 4.3.1 Confusion Matrices

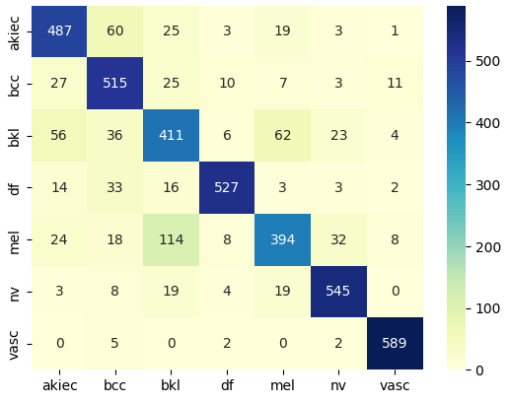


Figure 4.2: Customized CNN

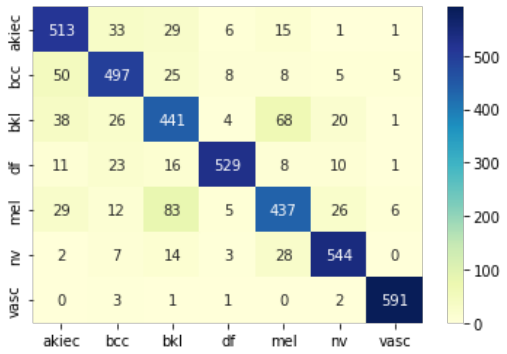


Figure 4.3: CNN-SVM

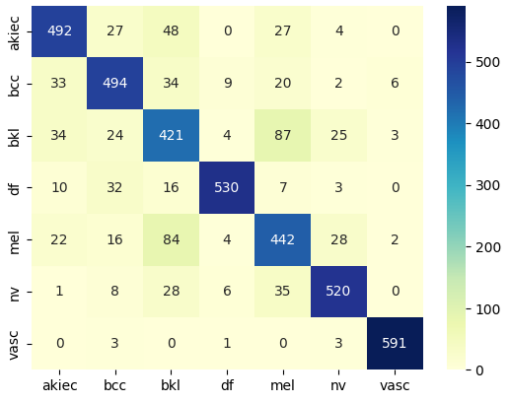


Figure 4.4: CNN-RandomForest

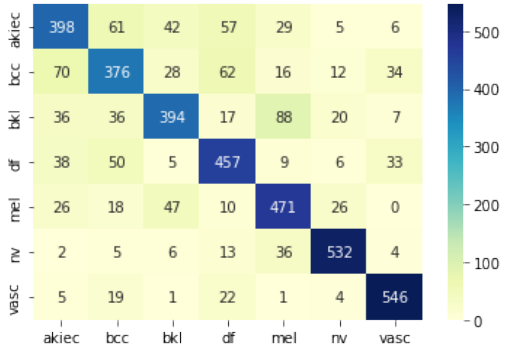


Figure 4.5: ResNet50

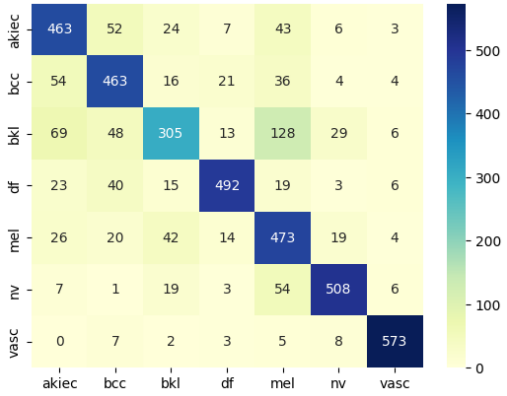


Figure 4.6: DenseNet121

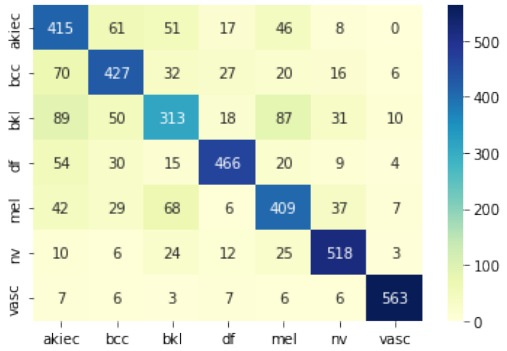


Figure 4.7: VGG19

This section outlines the precise and imprecise predictions of all models for each class of the dataset. A confusion matrix represents a summary of classification prediction outcomes. The number of exact and approximate predictions is calculated and broken down according to each class. Table 4.2 provides an overview of all ten confusion matrices.
Figure 4.2 demonstrates that the CNN-LSTM-SVM excellently performs accurate prediction.

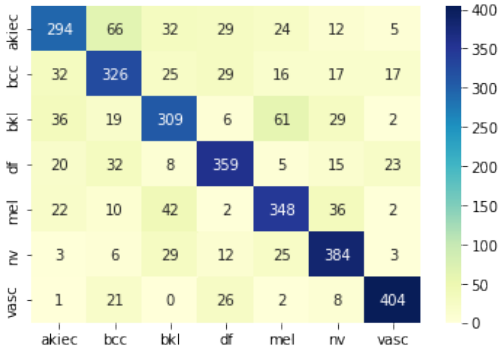


Figure 4.8: AlexNet

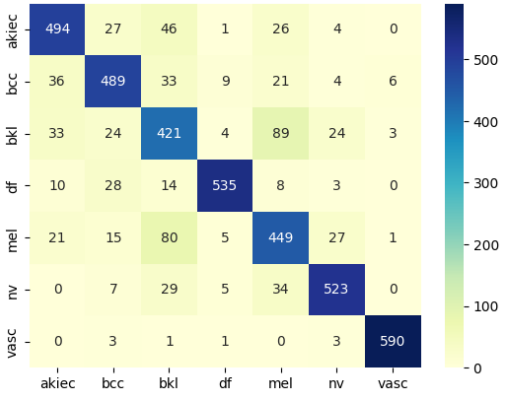


Figure 4.9: CNN-XGBoost

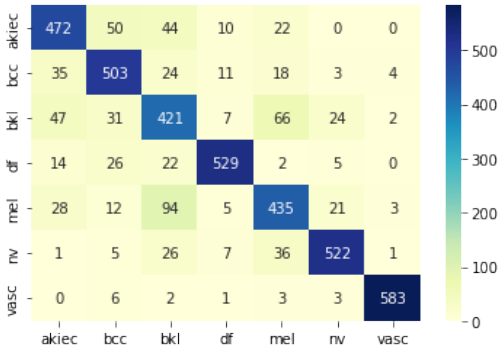


Figure 4.10: CNN-LSTM

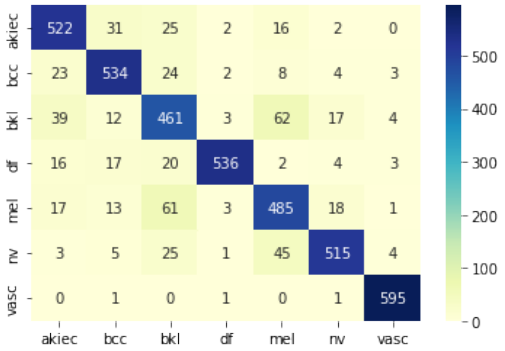


Figure 4.11: CNN-LSTM-SVM

### 4.3.2 Incorrect Prediction Graphs

For each exhibited model, incorrect prediction graphs were also plotted. The incorrect prediction graph depicts the percentage of wrong projections for each class. Through these ten graphs (Figure 4.12–Figure 4.21), it is evident that the estimation of CNN-LSTM-SVM is less than 25% inaccurate most of the time, but all of the other nine models have predictions that are between 30% and 50% inexact. The resulting accuracy and loss graphs of the CNN-LSTM-SVM model are outlined in Figure 4.22.

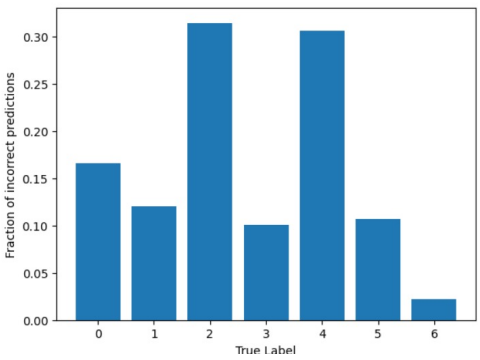


Figure 4.12: Customized CNN

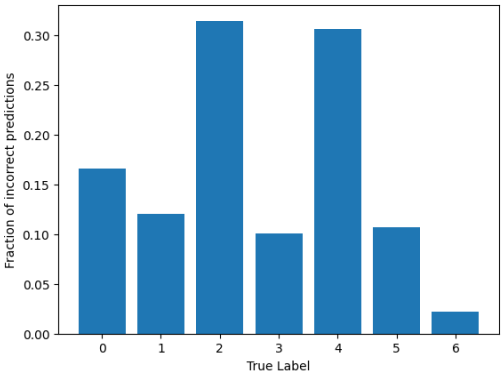


Figure 4.13: CNN-SVM

| Models | akiec | bcc | bkl | df | mel | nv | vasc |
|---|---|---|---|---|---|---|---|
| Customized CNN | 487 | 515 | 411 | 527 | 394 | 545 | 589 |
| CNN-SVM | 513 | 497 | 441 | 529 | 437 | 544 | 591 |
| CNN-RandomForest | 492 | 494 | 421 | 530 | 442 | 520 | 591 |
| ResNet50 | 398 | 376 | 394 | 457 | 471 | 532 | 546 |
| DenseNet121 | 463 | 463 | 305 | 492 | 473 | 508 | 573 |
| VGG19 | 415 | 427 | 313 | 466 | 409 | 518 | 563 |
| AlexNet | 294 | 326 | 309 | 359 | 348 | 384 | 404 |
| CNN-XGBoost | 494 | 489 | 421 | 535 | 449 | 523 | 590 |
| CNN-LSTM | 472 | 503 | 421 | 529 | 435 | 522 | 583 |
| CNN-LSTM-SVM | 522 | 534 | 461 | 536 | 485 | 515 | 595 |

Table 4.2: Summary of All the Confusion Matrices (598 Images for each class)

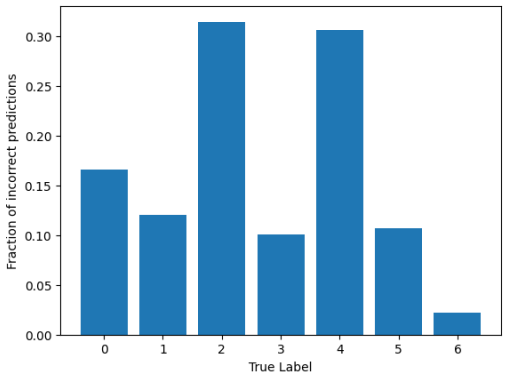


Figure 4.14: CNN-RandomForest

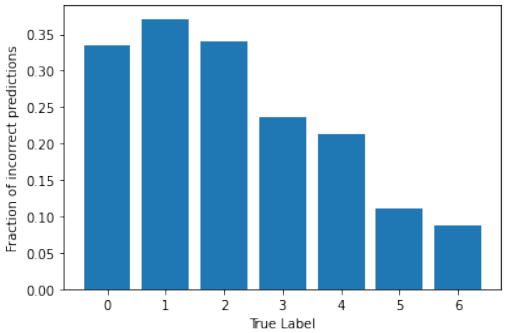


Figure 4.15: ResNet50

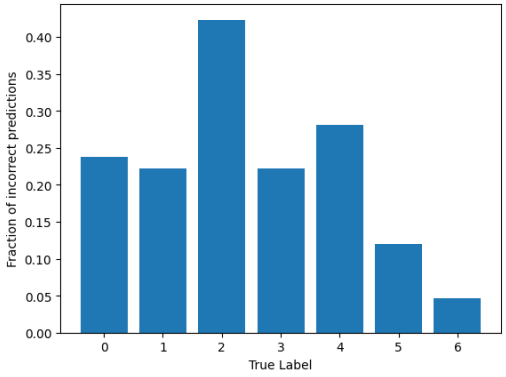


Figure 4.16: DenseNet121

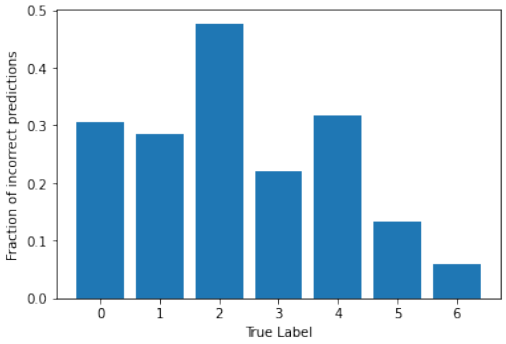


Figure 4.17: VGG19

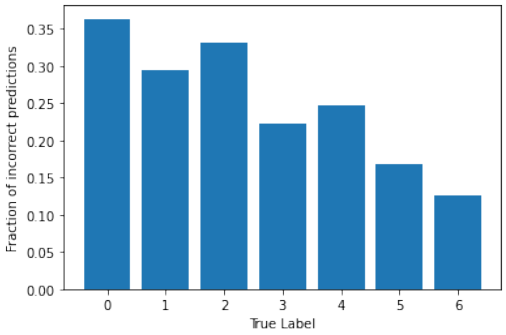


Figure 4.18: AlexNet

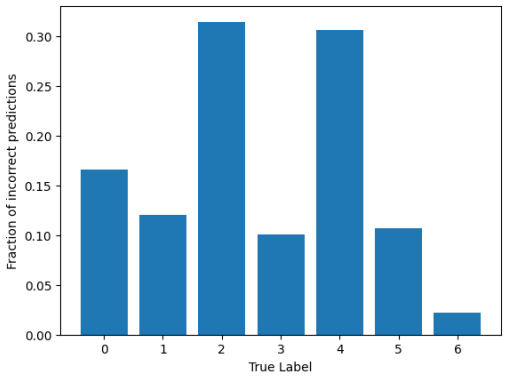


Figure 4.19: CNN-XGBoost

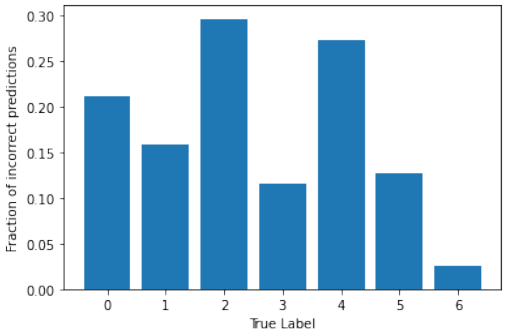


Figure 4.20: CNN-LSTM

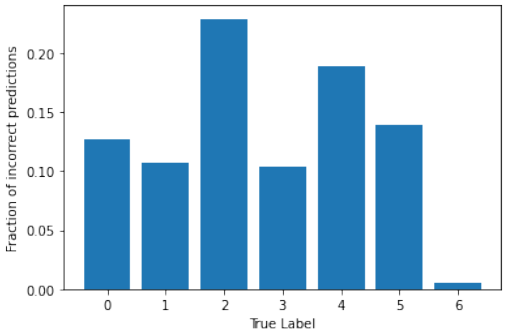


Figure 4.21: CNN-LSTM-SVM

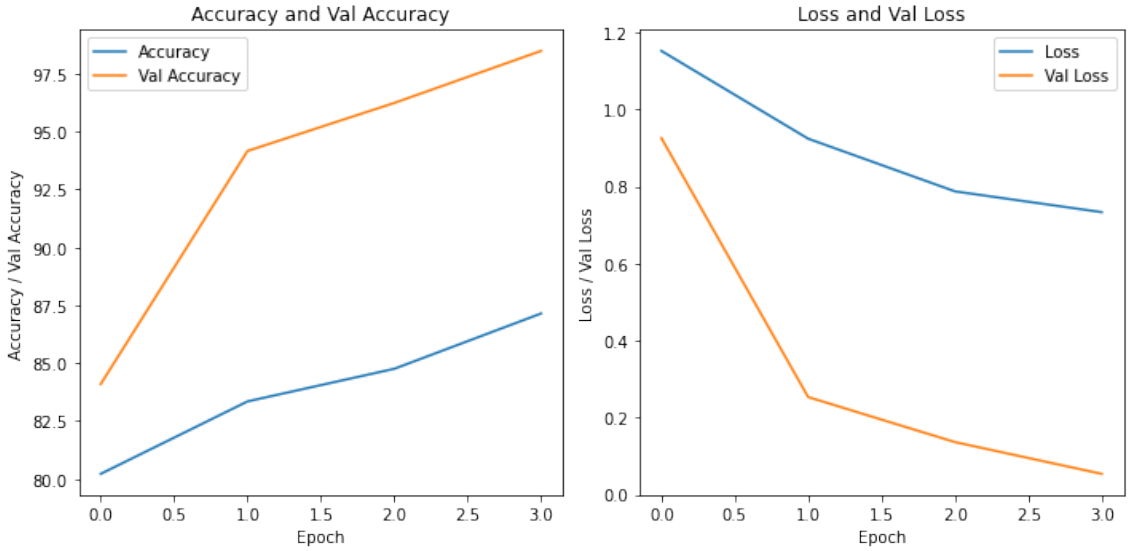


Figure 4.22: Accuracy and Loss Graphs of CNN-LSTM-SVM

# Chapter 5

# Future Endeavors and Conclusion

## 5.1 Future Endeavors and Conclusion

### 5.1.1 Future Endeavors

Due to the shortage of memory, we were unable to implement advanced preprocessing strategies. We preserved various forms of contrast images using the clahe, ahe, and fuzzy methods. Another filter known as morphologoyEx was also attempted to remove the hair and noise from images. In the future, we want to experiment with the images on the proposed model and analyze its performance. Additionally, we aspire to employ numerous datasets to demonstrate the outcome. Despite the popularity of HAM10000 data, the uneven distribution of data makes it inefficient. The study aims to perform comparative analysis on several multi-class cancer datasets.

### 5.1.2 Conclusion

In the past couple of years, there has been a dramatic increase in skin cancer cases. A report by the World Health Organization (WHO) estimates that skin cancer is the most prevalent cancer type, and is responsible for nearly 40 percent of all global cancer cases [8]. In recent papers, the performance of CNN and deep neural networks surpasses the accuracy of dermatologists.The accuracy of a dermatologist in detecting skin cancers ranges from 62% to 80%, whereas deep learning models can provide far superior results[12]. Still, multi-class detection is a complex task for both dermatologists and machine-learning models. To comprehend the small features of the classification process and determine the best model, the proposed models examine the performance of deep learning networks and self-created hybrid models. The customized CNN-LSTM-SVM exhibits outstanding performance in detecting the different classes of cancer. It is anticipated that a better pre-processing method will improve model performance even further. In brief, CNN-LSTM-SVM can be proposed as a suitable model for cancer detection due to its bidirectional feature learning properties.

# Bibliography


[1] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, pp. 273–297, 1995.

[2] D. Meyer and F. Wien, "Support vector machines," *R News*, vol. 1, no. 3, pp. 23–26, 2001.

[3] G. K. Kim, J. Q. Del Rosso, and S. Bellew, "Skin cancer in asians: Part 1: Nonmelanoma skin cancer," *The Journal of clinical and aesthetic dermatology*, vol. 2, no. 8, p. 39, 2009.

[4] F. F. Chamasemani and Y. P. Singh, "Multi-class support vector machine (svm) classifiers – an application in hypothyroid detection and classification," in *2011 Sixth International Conference on Bio-Inspired Computing: Theories and Applications*, 2011, pp. 351–356. DOI: 10.1109/BIC-TA.2011.51.

[5] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[6] T. Guo, J. Dong, H. Li, and Y. Gao, "Simple convolutional neural network on image classification," in *2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)*, IEEE, 2017, pp. 721–724.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[8] "Radiation: Ultraviolet radiation and skin cancer." Accessed on 2023-04-12. (2017), [Online]. Available: https://www.who.int/news-room/questions-and-answers/item/radiation-ultraviolet-(uv)-radiation-and-skin-cancer.

[9] M. Shaha and M. Pawar, "Transfer learning for image classification," in *2018 second international conference on electronics, communication and aerospace technology (ICECA)*, IEEE, 2018, pp. 656–660.

[10] R. C. Maron, M. Weichenthal, J. S. Utikal, *et al.*, "Systematic outperformance of 112 dermatologists in multiclass skin cancer image classification by convolutional neural networks," *European Journal of Cancer*, vol. 119, pp. 57–65, 2019.

[11] M. Vijayalakshmi, "Melanoma skin cancer detection using image processing and machine learning," *International Journal of Trend in Scientific Research and Development (IJTSRD)*, vol. 3, no. 4, pp. 780–784, 2019.

[12] S. S. Chaturvedi, J. V. Tembhurne, and T. Diwan, "A multi-class skin cancer classification using deep convolutional neural networks," *Multimedia Tools and Applications*, vol. 79, no. 39-40, pp. 28 477–28 498, 2020.

[13] K. E. Davis, *Skin cancer rates rising*, 2020.

[14] N. Hameed, A. M. Shabut, M. K. Ghosh, and M. A. Hossain, "Multi-class multi-level classification algorithm for skin lesions classification using machine learning techniques," *Expert Systems with Applications*, vol. 141, p. 112 961, 2020.

[15] F. W. Alsaade, T. H. Aldhyani, and M. H. Al-Adhaileh, "Developing a recognition system for diagnosing melanoma skin lesions using artificial intelligence algorithms," *Computational and mathematical methods in medicine*, vol. 2021, pp. 1–20, 2021.

[16] clevelandclinic. "Skin: Layers, structure and function." (2021), [Online]. Available: https://my.clevelandclinic.org/health/articles/10978-skin.

[17] M. Dildar, S. Akram, M. Irfan, *et al.*, "Skin cancer detection: A review using deep learning techniques," *International journal of environmental research and public health*, vol. 18, no. 10, p. 5479, 2021.

[18] R. C. Maron, J. G. Schlager, S. Haggenmüller, *et al.*, "A benchmark for neural network robustness in skin cancer classification," *European Journal of Cancer*, vol. 155, pp. 191–199, 2021.

[19] P. Aggarwal and F. A. Papay, "Artificial intelligence image recognition of melanoma and basal cell carcinoma in racially diverse populations," *Journal of Dermatological Treatment*, vol. 33, no. 4, pp. 2257–2262, 2022.

[20] K. Ali, Z. A. Shaikh, A. A. Khan, and A. A. Laghari, "Multiclass skin cancer classification using efficientnets–a first step towards preventing skin cancer," *Neuroscience Informatics*, vol. 2, no. 4, p. 100 034, 2022.

[21] M. Shafiq and Z. Gu, "Deep residual learning for image recognition: A survey," *Applied Sciences*, vol. 12, no. 18, p. 8972, 2022.

[22] *XGBoost - GeeksforGeeks — geeksforgeeks.org*, https://www.geeksforgeeks.org/xgboost, [Accessed 18-May-2023].