International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

SKIN CANCER CLASSIFICATION ON HAM DATASET by using CNN

P. Venkata Murlidhar [1], A.Sirisha [2], J.Srinidhi[3], K.Keerthika devi[4], Y.Joshanavi[5],A.HimaKalyan[6]

Department Of ECE

Aditya Institute Of Technology And Management,Kotturu, Tekkali

ABSTRACT:

The hallmark of skin cancer is the aberrant proliferation of skin cells, which is frequently brought on by UV radiation exposure from the sun or tanning beds. Basal cell carcinoma, squamous cell carcinoma, and melanoma are the three main forms of skin cancer. The most prevalent kind of cancer, basal cell carcinoma, derives from basal cells in the lower epidermis and typically appears on the head and neck as a result of sun exposure. Flat, scale-like squamous cells in the epidermis are the precursors of squamous cell carcinoma, which is also linked to sun exposure. The most aggressive kind of cancer starts from melanocytes and can spread to other areas of the body if it is not discovered quickly. In order to prevent and treat skin cancer, early detection and sun protection are essential. Because different hues appear identical, dermatologists have difficulty diagnosing skin cancer. Using the HAM10000 dataset, this study suggests a deep learning method for multi-class skin cancer classification. The created model uses convolutional neural networks to improve loss reduction and classification accuracy, which improves the diagnostic results for skin lesion categorization.

Introduction:

Made up of water, protein, lipids, and minerals, the skin is the body's largest and most important organ for protection. It covers the whole external surface and acts as a first-order physical barrier against the environment. Your skin controls body temperature and shields your body from pathogens. Sensations of heat and cold are sensed by nerves in the skin. The integumentary system includes your skin, hair, nails, sweat glands, and oil glands. "Integumentary" refers to the outer layer of a body. Its duties include controlling body temperature and guarding against pathogens, microbes, poisons, trauma, and ultraviolet (UV) light. Your skin controls body temperature and shields your body from pathogens.

Sensations of heat and cold are sensed by nerves in the skin. The integumentary system includes your skin, hair, nails, sweat glands, and oil glands. "Integumentary" refers to the outer layer of a body. Its duties include controlling body temperature and guarding against pathogens, microbes, poisons, trauma, and ultraviolet (UV) light. exposure, making it possible to identify it on a variety of skin areas. Additionally, skin that has been burned, harmed by chemicals, or exposed to x-rays may acquire it. Common sites for squamous cell carcinoma include the lips, skin surrounding the mouth, anus, and vagina. Melanoma.

There are three primary layers that make up the structure of skin:

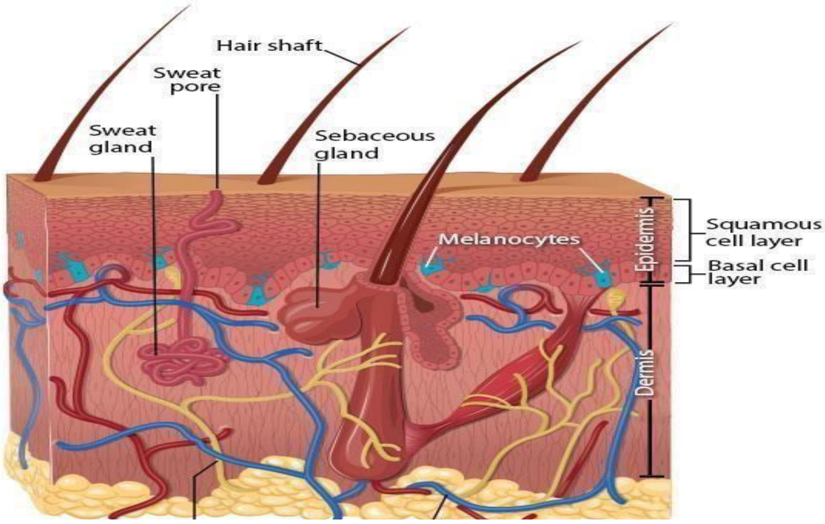
1. Epidermis: The skin's outermost layer acts as a barrier to protect. It is made up of multiple sublayers and is in charge of forming the skin's outer layer.
2. Dermis: The dermis, which is situated beneath the epidermis, is home to glands, blood vessels, nerves, and hair follicles. It gives the epidermis stability and sustenance.
3. Hypodermis (Subcutaneous Tissue): Fat and connective tissue make up the majority of this deepest layer. It acts as an energy storage system and aids in controlling body temperature.

   Three different types of cells comprise the epidermis.

- Squamous cells: The thin, flat cells that make up the epidermis' top layer.
- Basal cells: The squamous cells subordinate round cells.
- Melanocytes: Cells found in the dermal layer below the surface. The substance called melanin is what gives skin its hue. Skin darkens as a result of melanocytes producing more pigment when exposed to sunlight.

One kind of cancer that appears in the skin's outermost layer is called skin cancer. The main cause of skin cancer is the unchecked proliferation of aberrant skin cells as a result of UV radiation damage, which can occur from the sun or artificial tanning devices.

There are primarily three types of skin cancer.Melanoma,squamous cell carcinoma, and basal cell carcinoma.carcinoma with basal cells. The spherical cells in the bottom epidermis are called basal cells. This kind of cell is the source of about 80% of skin cancer cases.Basalcell carcinomas is the term used to characterize these tumors. Most frequently, basal cell carcinoma appears on the head and neck.



It generally results from sun exposure or appears in those who have radiation therapy as youngsters. They seldom spread to other body parts and typically grow slowly. carcinoma with squamous cells. Squamous cells are flat, scale-like cells that make up the majority of the epidermis.

About 20 percent of skin malignancies start out as these cells. Since sun exposure is the primary cause, it can be identified on a variety of skin places. Additionally, skin that has been burned, harmed by chemicals, or exposed to x-rays may acquire it. squamous cell On the lips and on the skin surrounding the mouth, anus, and vagina, carcinoma is frequently discovered. Melanoma. Where the epidermis and dermis meet, melanocytes are sporadic cells. The pigment melanin, which gives skin its color, is produced by these cells. The most dangerous kind of skin cancer, melanoma, begins in the melanocytes. It makes up around 1% of all cases of skin cancer.

TYPES OF SKIN CANCER:

There are various kinds of skin cancer, each with unique traits and possible effects on treatment and diagnosis. Here are specifics regarding the seven primary kinds:

1. Keratoses of Actinic
2. Cancer with Basal Cells
3. Benign Lesions Similar to Keratosis
4. Dermatofibroma
5. Melanoma
6. Nevi Melanocytic
7. Damage to the Vascular Organs

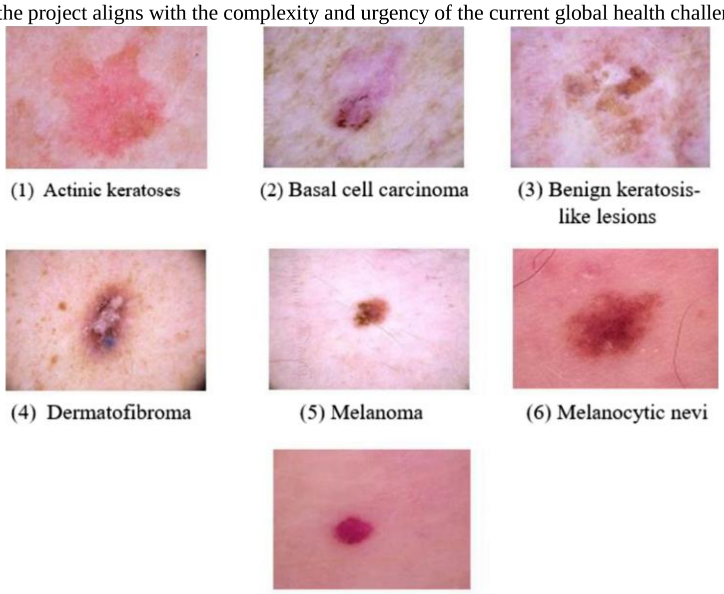
1. Keratoses of Actinic: Prolonged sun exposure can result in these precancerous growths. They are classified as squamous cell carcinoma in their early stages. Among its symptoms are Skin that has rough, scaly, or dry spots. usually found on parts of the face, ears, neck, and hands that are exposed to the sun.Patches can have flesh, pink, or crimson colors.
2. Cancer with Basal Cells :The most prevalent kind of skin cancer, BCC is frequently brought on by a lifetime of sun exposure. Among its symptoms are waxy or pearly lump with blood vessels clearly visible. a level lesion the color of flesh or
3. exposure. Among its symptoms are waxy or pearly lump with blood vessels clearly visible. Lesion that is flat and has the color of skin or a brown scar. can bleed or develop a crust.
4. Benign Lesions Similar to Keratosis: This group comprises a range of benign, non-cancerous lesions that may have characteristics of malignant growths. Lesions can look very different. Borrheic keratosis is a common form that manifests as a warty, stuck-on growth.
5. Dermatofibroma: a fibrous, benign nodule that frequently forms following mild skin damage. Among its symptoms are Reddish-brown, hard, little pimple on the skin. usually observed on the lower limbs.
6. Melanoma: Melanocytes, the cells that produce pigment, are the source of this potentially fatal skin cancer.

SCOPE AND OBJECTIVE:

SCOPE: The project's goal is to use Convolutional Neural Networks (CNNs) to build a deep learning model for automated skin cancer detection. The study primarily concentrates on the HAM10000 dataset, which has a wide range of pictures of skin lesions. The goal of the model is to categorize these photos into several skin cancer groups, offering a useful resource for early detection and treatment. This project's main objective is to painstakingly create a Convolutional Neural Network (CNN) model intended for the automatic detection of skin cancer by examining minute details in dermatological photos. This is a large-scale project with many moving parts, such as the methodical collection and preparation of various datasets, the tactical implementation and training of CNN architectures, and a critical assessment of the model's performance indicators.

OBJECTIVE: Building a reliable and efficient algorithm that can differentiate between dermatologist photographs of people with skin cancer and those without is the main goal of this project. This main objective informs all of the project's complex design decisions. In order to outperform existing benchmarks and get higher degrees of accuracy, the project purposefully chooses CNN models with a large number of layers, optimal top different activation algorithms. The thoughtful incorporation of intricate architectural elements guarantees the model's proficiency in the intricate process of recognizing the precise type and stage of skin cancer detection.

The vast and expansive nature of the project aligns with the complexity and urgency of the current global health challenges.



(1) Actinic keratoses (2) Basal cell carcinoma (3) Benign keratosis-like lesions

(4) Dermatofibroma (5) Melanoma (6) Melanocytic nevi

(7) Vascular lesions

EXISTING SYSTEMS:

LITERATURE SURVEY:

1. With a melanoma diagnosis accuracy of 85%, Vijayalakshmi M [2] created "Melanoma Skin Cancer Detection using Image Processing and Machine Learning," which investigates the use of Back Propagation Algorithm (Neural Networks), Support Vector Machine (SVM), and Convolutional Neural Networks (CNN). The stated accuracy may only apply to one model, hence it would be beneficial to compare the three algorithms' (BPNN, SVM, and CNN) respective performances.For training and assessment, the International Skin Imaging Collaboration (ISIC) dataset was used. The study by Vijayalakshmi M.M. investigates machine learning's potential for melanoma detection overall.
2. A Deep Convolutional Neural Network (CNN) approach improved with data augmentation for skin lesion classification was created in the paper "Deep CNN and Data Augmentation for Skin Lesion Classification" by Tri-Cong Pham, Chi-Mai Luong, Muriel Visani, andVan-Dung Hoang[4]. Improvements in AP (73.9% vs. 71.5%), Accuracy (87.2%), and AUC (89.2% vs. 87.4%) are reported by the study. The evaluation dataset is a publicly available skin lesion testing dataset.

INFERENCES:

We have identified key tactics that will guide the project's execution. First, because of their proven effectiveness in picture classification tasks, we are inclined to use Convolutional Neural Networks (CNNs) as the foundation of our model design. Second, we emphasize how important it is to fine-tune learning rate parameters in order to improve model training and convergence efficiency. We plan to take advantage of the Adam optimizer's known benefits from earlier studies in order to better enhance our training procedure.

We acknowledge that having too many techniques combined into one model might lead to complexity and difficulty in understanding. a large dataset that has been carefully selected to capture a wide range of skin cancer detection presentations. In order to minimize needless complications, we are streamlining our model by concentrating on a single, highly efficient CNN design. Finally, we're placing a lot of emphasis on using a sizable and carefully selected dataset that spans a variety of skin cancer types to train our model. Our model performs effectively under many conditions and is classified into distinct classes.

DESIGN AND METHODOLOGY:

PROPOSED SYSTEM:

The deep learning strategy with CNN model and Adam optimizer that is used in the suggested skin cancer detection methodology represents a major improvement over previous approaches, with the main objective being to increase accuracy levels. Notably, current systems display a wide range of architectural complexity, indicating a clear gap in attaining the breadth and depth typical of modern deep learning techniques. Because system designs differ so much, performance measures also differ. While some systems achieve remarkable accuracy rates, others face difficulties due to large dataset sizes, complex architectures, and optimization techniques.

DATA SET DESCRIPTION:

The HAM10000 dataset's major goal is to support research and development in the field of skin cancer diagnosis by utilizing deep learning and machine learning methods [8]. The photos are dermoscopy pictures that were taken by different medical specialists, guaranteeing a varied and representative assortment of skin blemishes. With 10,015 dermoscopy images, the collection is among the biggest and most complete for research on skin cancer. Dermatopathologists have contributed comprehensive annotations to each image in the dataset, which show the skin lesion's ground truth label.

DATA SET PREPROCESSING:

Preparing the HAM10000 dataset for your skin cancer detection project entails a number of crucial actions that are necessary to get the data ready for deep learning mode training. following Using your preferred programming environment, load and examine the dataset to learn about its structure, look for any missing values, and obtain understanding of how different classes are distributed.

DATA SET AGUMENTATION:

Applying different changes to preexisting photos to produce new, slightly altered ones is known as data augmentation. By using this method, the dataset's effective size is increased and the model is given a wider range of samples to learn about. The answer is to generate extra data through data augmentation in order to remedy the insufficient dataset size. Image data augmentation methods include random rotation, flipping, zooming, and adjusting brightness and contrast.

DATA SET COMPILATION:

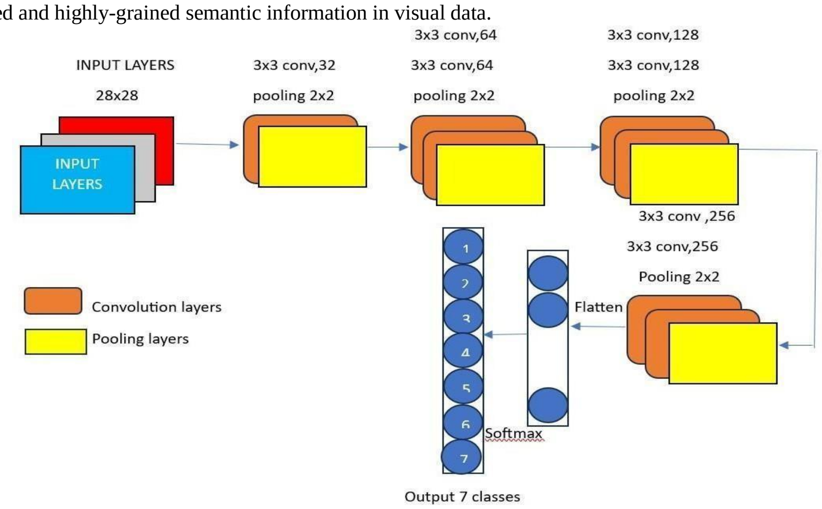
When referring to the process of organizing and getting ready the data for your machine learning model's training, validation, and testing stages, you are referring to the HAM10000 dataset in the context of skin cancer detection. It involves duties like such as dividing the dataset into smaller groups, adding labels and pictures, and preparing the data in a way that will help your deep learning model be trained.

MODEL ARCHITECTURE:

With the aim of improving robust pattern recognition and optimizing accuracy, the proposed Convolutional Neural Network (CNN) architecture for the skin cancer detection model using the HAM10000 dataset has a carefully constructed design. This CNN creates a fully linked segment specifically for classification by combining many convolutional layers with pooling layers in a seamless manner. Convolutional layers: these layers are good at extracting hierarchical characteristics from input images; pooling layers: these layers systematically lower spatial dimensions, which helps the model extract important information. Based on whether or not malignant cells are present in a skin lesion image, the model is intended to automatically classify those images into various categories. With the help of training and validation loss and accuracy metrics, the model gains the ability to identify pertinent patterns in the input photos and generate precise predictions.

CNN ARCHITECTURE AND LAYERS:

One kind of neural network architecture that excels at processing and categorizing visual data—such as photos, movies, etc.—is the convolutional neural network (CNN). Because CNNs can automatically learn from and extract features from photos of skin lesions, they are effective. The primary applications of convolutional neural networks (CNNs) are in image recognition and classification. It is useful for a number of tasks in computer vision, image processing, object recognition, and face recognition. They are efficient at automatically picking up hierarchical feature representations and identifying both fine-grained and highly-grained semantic information in visual data.



1.Convolutional layer
2.Max pooling layer
3.Batch normalization layer
4.Flatten layer
5.Input layers
6.Output layers
7.Softmax function

Layer name	Dimension (H,w)	Filter	Parameters
conv2d (Conv2D)	(28,28)	32	896
Leaky relu	(28,28)	32	0
max_pooling2d	(14,14)	32	0
batch normalization	(14,14)	32	128
conv2d_1	(14,14)	64	18496
Leaky relu _1	(14,14)	64	0
conv2d_2	(14,14)	64	36928
Leaky relu _2	(14,14)	64	0
max_pooling2d _1	(7,7)	64	0
batch normalization _1	(7,7)	64	256
conv2d_3	(7,7)	128	73856
Leaky relu _3	(7,7)	128	0
conv2d_4	(7,7)	128	147584
Leaky relu _4	(7,7)	128	0
max_pooling2d _2	(3,3)	128	0
batch normalization _2	(3,3)	128	512
conv2d_4	(3,3)	256	295168
Leaky relu _4	(3,3)	256	0
conv2d_5	(3,3)	256	590080
Leaky relu _5	(3,3)	256	0
max_pooling2d _3	(1,1)	256	0
batch normalization _3	(1,1)	256	1024

Table 3.1. Tabular form of layers

OPTIMIZER AND ACTIVATIONFUNCTION:

OPTIMIZER: Deep neural network training frequently uses the well-liked optimization algorithm known as the Adam optimizer. This algorithm is for adaptive learning rate optimization. Deep neural network training frequently makes use of the adaptive learning rate optimization technique known as Adamoptimizer. It incorporates concepts from Momentum and Root Mean Square Propagation (RMSProp), two additional well-known optimization methods. For every parameter, Adam keeps two exponentially declining averages.

The gradients' first moment (mean) and second moment (uncentered variance)The initial moment aids the optimizer in monitoring the gradients' progress and is comparable to the momentum term in stochastic gradient descent (SGD).The scaling of the learning rates for every parameter is associated with the second moment.

This code's Adam optimizer:

θt+1=θt - learning rate is the formula for the Adam optimizer. rootover(v^t)+ϵ + xm^t

The hyperparameter for learning rate is θt, which is the parameter at time step t.

First moment estimate corrected for bias is denoted by m^t. The estimate of the bias-corrected second moment is v^t. For numerical stability, ρ is a tiny constant.

The learning rate has been set to 0.004, while the default values of the other hyperparameters (such as 1β, 2β2, and ϵ) are maintained. These default values are commonly utilized in the Adam optimizer.

ACTIVATION FUNCTION:

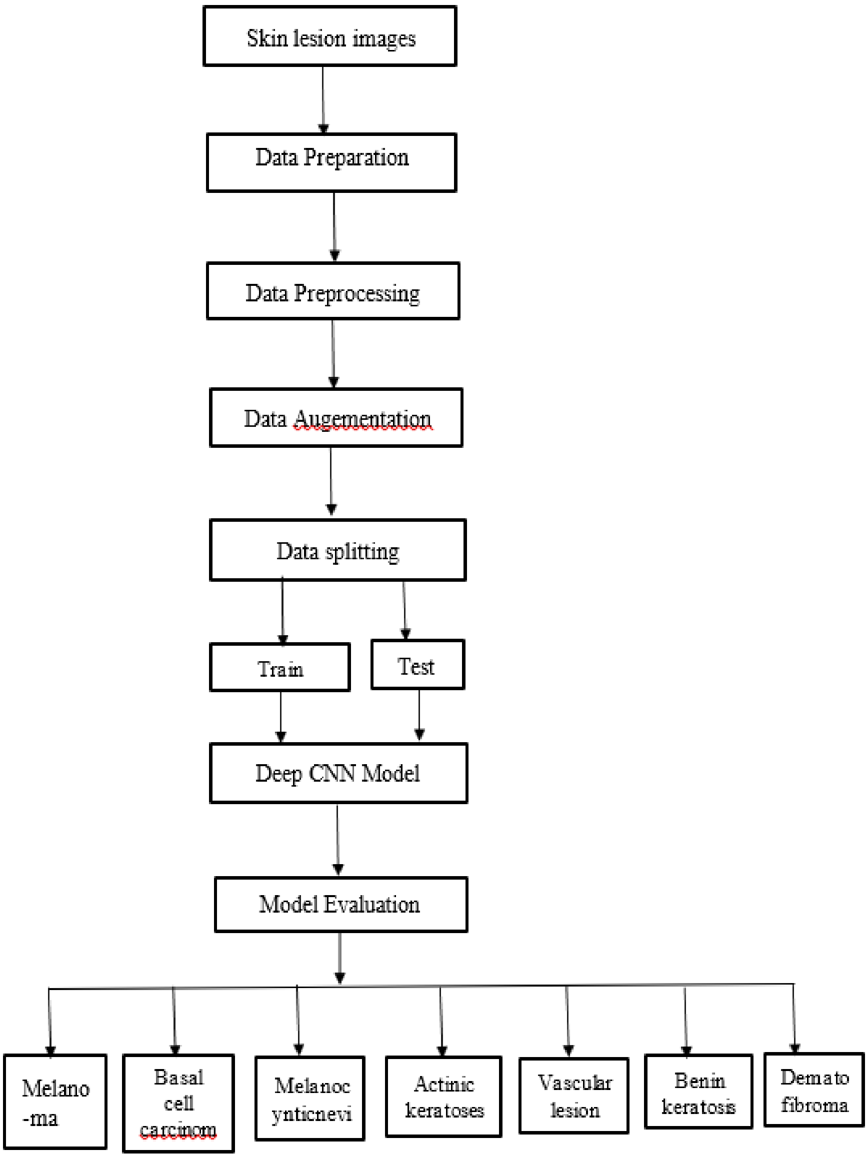
The Leaky ReLU function in our project code will output 0.2 times the input value for negative inputs, allowing a tiny gradient to pass through the network during training and assisting in mitigating the "dying ReLU" issue. Leaky ReLU is a popular option for activation functions in different neural network designs because it strikes a reasonable mix between the ease of use and efficiency of ReLU and its capacity to accept negative inputs.gives negative inputs a tiny positive slope that makes their impact on activation negligible.

In terms of math, F(x)=max(a*x,x)

In this case, a,(alpha) is a tiny constant that usually ranges from 0.01 to 0.1. By guaranteeing that even negative inputs make a tiny contribution to the output, this slope avoids "dying ReLU" and enables the network to learn from a larger variety of inputs.

1. High predictive accuracy
2. Handles nonlinear relationships
3. Robust to outliers

METHODOLOGY:



IMPLEMENTATION AND RESULT:

SOFTWARE ENVIRONMENT:

Software package	Purpose
Python	Base programming language.
Open CV	Provide tools for image and videoprocessing
Keras	High-level deep learning API for building and training models.
TensorFlow	Backend used by Keras (may not be needed if your Keras installation includes it automatically).
NumPy	Fundamental library for numerical computations.
Matplot lib	Main plotting library for creating graphs.
Sea born	It is a statistical data visualization library based on Matplotlib

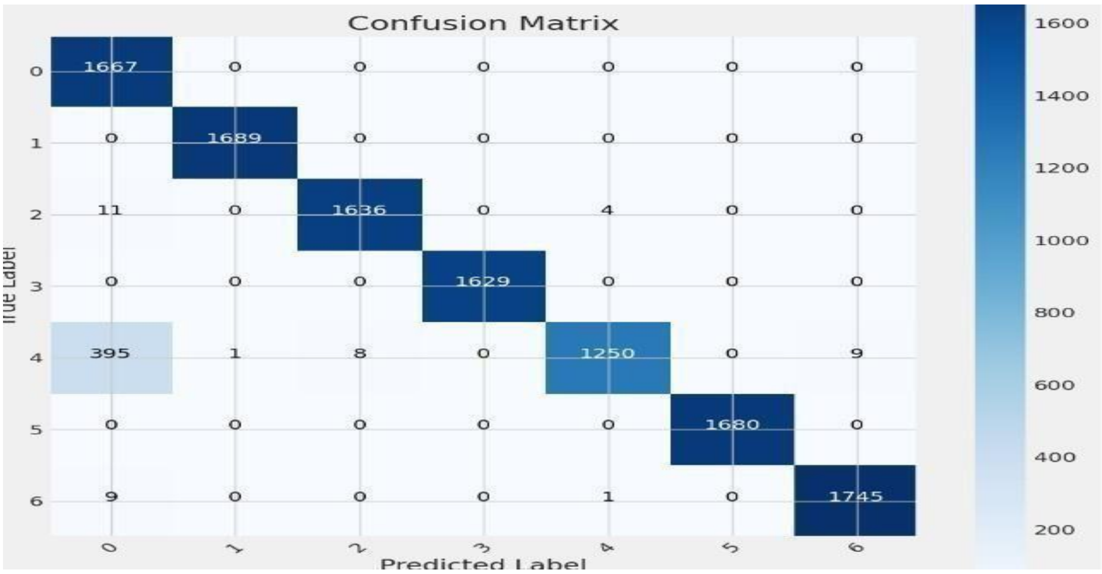
RESULTS:

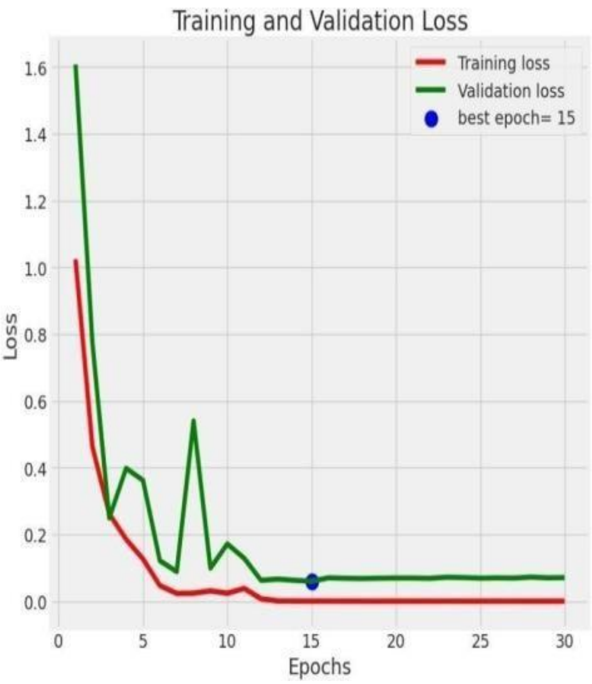
```
Class 2 - Total: 1099, Train: 824, Test: 275
Class 4 - Total: 6705, Train: 5028, Test: 1677
Class 3 - Total: 115, Train: 86, Test: 29
Class 6 - Total: 1113, Train: 834, Test: 279
Class 5 - Total: 142, Train: 106, Test: 36
Class 1 - Total: 514, Train: 385, Test: 129
Class 0 - Total: 327, Train: 245, Test: 82
```

```
Classification Report:
              precision    recall  f1-score   support

           0       0.80      1.00      0.89      1667
           1       1.00      1.00      1.00      1689
           2       1.00      0.99      0.99      1651
           3       1.00      1.00      1.00      1629
           4       1.00      0.75      0.86      1663
           5       1.00      1.00      1.00      1680
           6       0.99      0.99      0.99      1755

    accuracy                           0.96     11734
   macro avg       0.97      0.96      0.96     11734
weighted avg       0.97      0.96      0.96     11734
```





CONCLUSION:

"Skin lesion classification on HAM 10000 dataset by using CNN" demonstrates the effective deep learning techniques, particularly convolutional neural networks (CNNs), accurately detecting skin cancer from dermatoscopic images. By leveraging the HAM10000 dataset, which comprises a diverse collection of skin lesion images, we have developed a robust and reliable model capableof distinguishing skin cancers with high accuracy of 96%.

Through meticulous experimentation and model optimization, we have achieved promising In cause terms of classification accuracy and performance metrics. The CNN architecture, with itsconvolutional layers, max-pooling layers, and fully connected layers, and activation function leakyReLu in order to nonlinearity and Adam optimizer .

REFERENCES:

1. https://www.researchgate.net/publication/220427986_The_Appearance_of_Human_Skin
2. Vijayalakshmi M M "Melanoma Skin Cancer Detection using Image Processing and Machine Learning" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-3 | Issue-4, June 2019, pp.780-784.
3. Lakshay Bajaj, Himanshu Kumar, Yasha Hasija . Automated System for Prediction of Skin Disease using Image Processing and Machine Learning. International Journal of ComputerApplications. 180, 19.