PROJECT REPORT

TOPIC- SKIN CANCER DETECTION USING MACHINE LEARNING AND DEEP LEARNING

Submitted by:

Krishanu Bakshi (M.sc, Statistics)

Under the guidance of:

Prof. Susmita Ghosh

CSE dept (Jadavpur University)

Content

❖ Abstract:

Skin cancer is a common type cancer. In every year millions of people are affect by this cancer. Three million people are affected in USA by this cancer every year. The rate of survive ness also decreases as this diseases progress. However, detecting of skin cancer in early is very difficult for dermatologist that's why people are using supervised and unsupervised technique and also use deep learning technique to detect skin cancer.

A novel frame work based on deep learning and machine learning are proposed in this study for multiclassification skin cancer types such as melanoma, dermato-fibroma, pigmented benign keratosis. nevus, basal cell carcinoma, vascular lesion. The proposed model is XG Boost model. Moreover, the accuracy of the proposed method also compared with the two bosting technique which are nothing but Ada boost & Cat Boost also compare with Decision tree, KNN, Support vector machine and one deep learning model that is multi-layer perceptron model. This comparative analysis allows for a comparative assessment of proposed model's performance.

Data pre-processing is a very important role to enhancing model performance.

In this paper 2017's skin cancer data sets were used, where different types of dermoscopic images represent various types of skin cancer. Images were resized, normalized and extracted features to improve model performance. Advance technique like Synthetic Minority Oversampling Technique also used for class imbalances and improve the model performances. Additionally, feature was extracted using vgg16 which is a pretrained convolution neural network.

The experimental result indicates that XG boost model perform better than others and it achieved 98.89% test accuracy. This suggest that the gradient boosting model like XG boost were situated for classification task. In contrast the other model following the accurses: 77% accuracy found from multi-layer perceptron,

Decision tree at 95.08%, KNN at 55.13%, SVM at 67.63%, Ada boost at 32.75

% and Cat boost at 95.25 %.

❖ Motivation and Contribution-

The research goal is to create a model which predict the image is melanoma or not and identify the skin lesions accurately an early stage. It will reduce the mortality rate. The key issue is maintaining in following manner.

Early Detection of skin cancer is very important for successful treatment outcome.

Machine learning and Deep learning techniques offer potential to enhance diagnostic accuracy, even in early stage by automating and improving feature extenuation and classification.

The increase of skin cancer patient in globally is really high so to prevent the early detection to reduce the mortality rate machine learning based frame works can provides diagnostic support to dermatologist allowing them diagnosed the disease quickly and give them more consistent evaluation particularly in the imitated region of health care.

By creating multiclass classification model this study can determine several types of skin cancer like melanoma, basal cell carcinoma etc.

Self – diagnosis may be takes lot more times for a dermatologist by judgment and experience. By providing an automated, standardized result This method can enhance subjectivity and contributing to great consistency in skin cancer diagnosis.

This study is not only a novel deep learning frame work but extensive comparative analysis with few machine learning models.

The complex visual feature is actually very difficult to capture accurately. By using deep learning this study enhance feature extraction capability, allowing the model to learn complex pattern and characteristic that may be difficult in manually.

This research used 2017 skin cancer data to train and validate the model. Testing on this dataset ensure that model performance is robust and relevant for real world clinical application.

❖ Literature Review-

This paper [1] found that the MobileNet V2 architecture fit the ISIC 2020 skin cancer dataset really well. The model achieved 98.2 % average accuracy. The transfer learning model obtain 98.3% & 98.0 recall on melanoma and benign and also found that the f1 score was 98.1%. In the experiment process the Adam optimizer, binary-cross-entropy loss function, 100 epochs, 64 batch size, and default alpha rate were used.
From confusion matrix this paper found that the MobileNetV2 model correctly identify 1721 benign images lesion out of 1750 and 1556 malignant out of 1750.

This paper [3] proposed multiclass support vector machine model. This paper used ISIC 2019 data sets. In this paper, Dull Razor method and Gaussian filter are used for image enhancement and median filter is used for noise removal. The above steps are considered as preprocessing steps. Color-based-k-means clustering used to segmented images.
Here multiclass support vector machine achieved 96.25% accuracy.

This paper [8] proposed a deep convolutional neural network for skin cancer identification from dermoscopy images. In this paper they were used HAM10000 data set. The results indicate the strength of deep learning in the skin cancer detection process. Here, the deep learning model achieved 84% accuracy. The ROC curve shows the trade-off between sensitivity and specificity. The model achieved 0.91 AUC with a threshold of 0.5.

This paper [13] found that DCNN model is best transfer learning model to classify whether a skin lesion is malignant or benign. The ability of proposed method is to classify benign or malignant lesion by replacing output activation layer with sigmoid for the binary classification. Moreover, the proposed method was evaluating on a data set named HAM10000. In this study this paper proposed DCNN model in two ways with 70% training images and another is aimed to evaluate the proposed model with 80% of training images that shows best accuracy. Although some other transfer learning models like Alexnet, VGG16, ResNet, DenseNet, and MobileNet also used for the same data set but DCNN model perform really well over the same data set. In this paper the DCNN model performed with 200 epochs but on the training process the model got overfitted. The best results found in 100 epochs and obtained AUC score 0.847 also the model got 91.43 testing accuracy which was better result than the other models. Also, this paper suggests

that for early detection of malignancy in skin cancer especially for those without admittance to doctor can significantly encourage them to get the treatment and enrich the survival possibility.

This paper [28] experiments on the ISIC 2019 data set. This paper found that Melanoma cancer is a most dangerous skin cancer to detect early this cancer this paper uses Mobile net model. The proposed model implements on the ISIC 2019 data set. CNN demonstrated a classification accuracy of 0.86% but the proposed model was the smart Mobile net and it achieved 0.89% accuracy indicate the significant number of correctly identified instance. The Precision, sensitivity and F-measure all exhibit the value of 90%, which significant a high level of accuracy and correctly detecting actual positive instance. In general, Smart mobile net model exhibits a robust performance in the identification of skin cancer.

This paper [29] used the HAM10000 data set for their experiments. In the pre-processing steps they normalize the skin image, reduction the data, and augmented the data to classify malignant or benign from the relevant dataset. This paper uses Alex net, Res net, Inception V4 & VGG16 model. But their proposed model achieved highest train and test accuracy which were 96.1 & 90.93 also the precision, recall and F1 score were 94.69,94.13 & 94.73 which were very high. From those high result this paper also clarify that the model mostly correctly identified the positive classes.

This paper [17] used ISIC data set for the experiments. After the pre-processing steps they used 3-layer CNN, VGG16 & Google Inception V3 and found that Google inception V3 model fitted well. It achieved 90% training & 81% testing accuracy and the sensitivity was 84.33%. With these high scores, deploying CNN with Google Inception V3 will complement the efforts of dermatologists for diagnosing the skin cancer.

This paper [30] used ISIC 2019 data set for the experiments. This paper used various type of transfer learning algorithms like Resnet 50, Inception v3, AlexNet and VGG19. Also, this paper proposed a new model which is called SCDNet classifier which is combine with VGG16 & CNN for the multi-class classification problem. The accuracy of the SCDNet was 96.91% which was higher than rest of using models. This paper also notice in this universe the melanoma caner increase and the other skin cancer increase rapidly. A speedy recovery needed for the large number of cases that's why deep learning model came in the context.

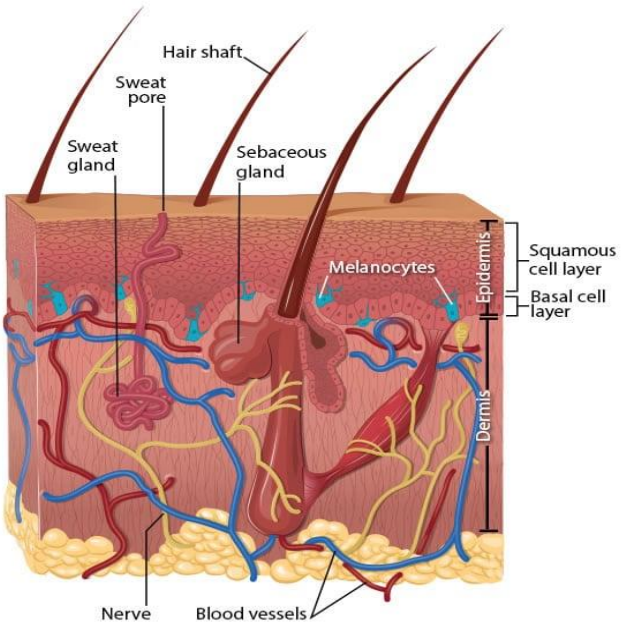
❖ Introduction:

The skin is a body's largest organ. It protects our body against infection and injury and helps to regulate our body temperature. It's also stores water & fat and produce vitamin-D. The skin has several layers but the main layers are Epidermis, Dermis [1]. Skin cancer is caused by uncontrolled growth of abnormal skin cancer cells which result in malignant tumor. When these cells are exposed in ultraviolet rays, a mutation occurs in the DNA which affects the normal growth and eventually in skin cancer [2]. Skin cancer accounts for one third of all cancer cases worldwide [3]. Dermoscopy is one of the most common technique to detect the skin cancer. The average accuracy of dermatologist is between 60% to 80% for skin cancer diagnosis using dermoscopic image. For the skin cancer classification, the hand-crafted features such as shape, colour and texture have been widely investigated. According to the American cancer society melanoma skin cancer cases are only 1% of total cases but the death rate is very high [6]. The American cancer society declined that Melanoma's death rate is increased rapidly during 2013 to 2017. Rates fell by 6% to 7% per year [7]. In 2015 in a survey there was found that 17.5 million cancer cases are world wide where 8.7 million people are died due to cancer [8]. Every year more than 5 million new case of skin cancer are registered in the United States [9]. In USA every year 10,000 deaths occurs due to melanoma [10]. In 2021, 106,110 cases of melanoma were registered where 7180 are died due to this cancer [11]. In 2022 197,700 new melanoma cases are arrived in USA [12]. Every year approx. 1000,000 new melanoma cases are detected in Europe [13]. In Australia 15,229 peoples are diagnosed by melanoma. Since 1990, skin cancer is increased in UK by 120% & in USA 255% increased. However, melanoma is a highly affected cancer if it is diagnosed at the initial stage. The survive rate is 96% in the initial stage, but it is dropped to 5% in advance stage.

In the pre-processing steps firstly resized and normalized the data set then used Histogram oriented gradient, Local binary patterns, Deep feature extraction using VGG16 & colour feature extraction. Then concatenate those feature and train test split into 80-20 ratio also labelling the train and test label using the label encoder so that the model can understand the label's really well. At last use Some machine learning model and deep learning model and found that XG boost model achieved highest accuracy and higher precision, recall results. It means that XG boost model correctly classify the most of the positive classes.

The train accuracy and the test accuracy were 95.47% & 98.89%.

❖ Essential Definition-



Skin Cancer begins in the Epidermis layer which made up of three kinds of cell –

a) **Squamous cells-** Thin, fat cell that from the top layer of the epidermis.

b) **Basal cells** –Round cells under the squamous cell.

c) **Melanocytes-**This cells found in the lower part of the epidermis

Cancer begins when healthy cells change and grow out of control, forming mass called tumor A tumor can be cancerous or benign. A cancerous tumor is called malignant it means it can grow and spread to the other body parts. A cancerous tumor is benign tumor it means it can grow but not spread to the other part of the body.

Here Basal and squamous carcinoma is two most common type of skin cancer. They begin in basal and squamous layers of the skin respectively. Melanoma is the third most common type of skin cancer begins with melanocytes. Melanoma causes the most deaths because it's tendency to spread to the other body parts.

Here I briefly talked about some skin cancer –

a) **Actinic keratosis-** An actinic keratosis is a rough, scaly patch on the skin that develops from years of sun exposure. It's often found on the face, lips, forearms, scalp, neck or back of the hand. This is also known as solar keratosis. An actinic keratosis grows slowly and usually first appears in people over 40 years.

b) **Basal cell carcinoma-** This is a most common type of cancer. Basal cell carcinoma begins with basal cells. This is a type of cell that creat new skin cells as old cells dies.

This is often appears as a slightly transparent bump on the skin, though it can take other from. It occurs most often where the skin is exposed to the sun such as head or neck.

Mostly Basal cell carcinoma occur from ultra violet radiation from the sunlight.

c) **Dermatofibroma-** It is the benign nodule in the skin. It is found in elbows, chest or neck. It is usually pain less The size of this tumor is basically 0.2 cm to 2 cm. From mild trauma like insect bites.

d) **Melanoma-** Melanoma which means black tumor is the most dangerous type of skin cancer. It's grows quickly and has ability to spread to any organ. Melanoma comes from skin cells called melanocytes. These cells produce melanin, the dark pigment that gives skin its color. Most melanomas are black or brown in color, but some are pink, red, purple or skin colored. Melanoma accounts for only about 1% of all skin cancers, but causes the great majority of skin cancer-related death. It's one of common cancer in young people under 30.
It is dramatically increased past 30 years. It's widely accepted that increasing labels of ultraviolet exposure are one of the main reasons for the rapid rise in the number of melanoma cases.

e) **Nevus** – It is a medical term for a mole, a common type of skin growth. Nevi refers to multiple moles. It is also called dysplastic nevi for multiple atypical moles.

f) **Seborrheic keratosis-** A seborrheic keratosis is a common benign growth, similar to a mole. Skin growth like seborrheic keratosis are sometime also called epidermal tumor. The tumors appear in various color from light tan to black and less often white, yellow or pink. They are round or oval feel flat or slightly elevated. Its size is very small to more than 2.5 centimeter across.

g) Squamous cell carcinoma-

This type of skin cancer actually happened for the growth of cells on the skin. It is a common type of skin cancer. This kind of cancer found in firstly in the Squamous cell.
Mostly Squamous cell carcinoma happened for too much ultraviolet radiation. It can be found anywhere on the skin. In people with black and brown skin, squamous cell carcinomas are more likely to be on skin that isn't exposed to sun.

h) vascular lesion-

A vascular lesion is an abnormal growth in the blood vessels that can occur in various part of body. They can be present at birth or acquired later in life. It can find as skin marks, tumor. It can be found on the surface of skin just below or deep in the vein tissue.

It can be happened from injury, infection, or other underlying medical condition.

❖ Data Description-

- **Data acquisition -**

The data set is found from the website Kaggle. The data set is bio medical related. In this data set there total 2357 images were there where the train, folder contain 2239 images and, in the test, folder contain 118 images.

❖ Methodology-

a) Data preprocessing-

It is very important steps in the data analysis pipe line that involves cleaning, transforming, and organizing raw data into suitable format of the data analysis
process. The goal of pre-processing is to ensure that the data is accurate, consistent and ready for the modeling and analysis.

b) Encoding –

In machine learning pipeline models are not understand the categorical labels. So, that's why generally transform the categorical labels into numerical labels. To transform categorical label to numerical label here used label encoding. In label encoding each label assigns a numerical value.

c) Feature extraction-

In image processing feature extraction is a most necessary step to extract the features from the image data sets. Histogram oriented gradient, Local binary pattern, color histogram and vgg16 model is use for feature extraction.

❖ Proposed Methodology-

1) Decision Tree-

A decision tree is a tree based supervised classification model. It is useful for Classification and regression problem. The decision tree algorithm recursively splits the data based on features to creat a tree that makes prediction.

a) Tree construction-

❖ **Root node-**

The algorithms start with a root node that contains the entire dataset.

It evaluates different features to find the one of the best splits the data set into distinct class. (e.g., Gini impurity or entropy for classification)

❖ **Splitting Nodes-**

One of the root nodes is split based on the selected feature, it creates the child nodes representing the different branches of the split. This process works recursively, with each node splitting the data based on another feature until a stopping criterion is met.

❖ **Leaf Nodes-**

When a node cannot split further or meets the stopping criterion it becomes the leaf nodes.

Leaf nodes represents the class label (classification) or average value of the target value.

b) Splitting Criterion-

- **Classification-**

for classification task common splitting criteria include Gini impurity & entropy.

Gini impurity measure the probability of misclassifying a randomly chosen element

If it were labeled randomly according to the distribution label of the node.

Entropy measures the uncertainty in a set of data.

The attribute with the highest information gain or lowest Gini index is selected as the split criterion.

- **Entropy-**

  The measure of impurity is called entropy. For 2 class problem the minimum entropy is 0 and the max entropy can be greater then 1.

  The formula of Entropy is-

  $E(s)=\sum_{i=1}^{n} -p_i * \log_2 (p_i)$

  Where p_i is the probability of an elements in the data

- **Information Gain-**

  Information Gain is a metric used to train decision trees. Spcially this metrics measure the quality of a split. The decision tree is all about finding attribute that returns highest information gain.

  Information Gain = Entropy(parent)-weighted average entropy(children)

- **Gini Impurity-**

  It is a measure of the likelihood of incorrectly classification of a new instance of data , if that new instance were randomly classified according to the distribution of class labels from the data set. So, it helps us to identify which feature is better to split.

  Gini impurity=$1-\sum_{i=1}^{c} p_i^2$

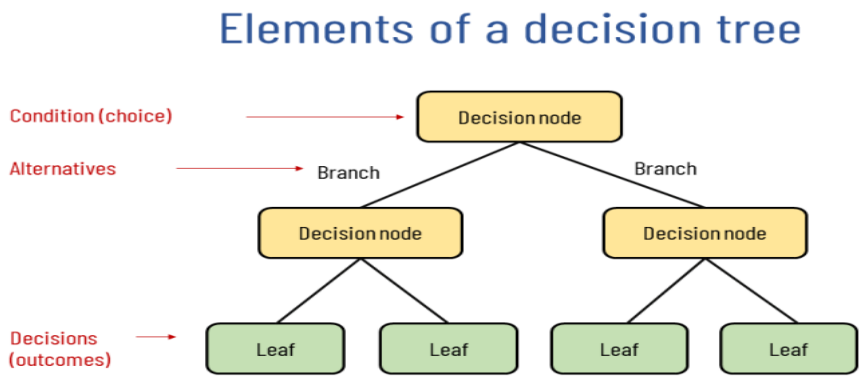


Figure 1: Decision Tree Workflow

2) KNN-

This algorithm known as Kth nearest neighbor. The algorithm classifies new data points based on the majority class of their k nearest neighbor. The algorithms store all the training data points and their corresponding labels. For The new data points the algorithms calculate the distance from the all training data points. It then select the k nearest neighbor. It is also called lazy learning algorithm because in the training phase it does not learn the data pattern but in the testing phase the model predict from the given data that's why it is called a lazy learning algorithm. It is also called instance based or memory-based learning algorithm because it memorizes the entire training phase and when a prediction is need it looks up the most similar instances in the training data. Key parameters of knn are K: The number of neighbors to consider. A larger K smoothness the decision boundary but can lead to mis-classification of points near the boundary , Distance matrix: To calculate the closest distance here typically used Euclidean distance, but other metrics like Manhattan, Minkowski distance are also use based on the nature of the data set.

a) The formula of Euclidean distance is – $d = \sqrt{\sum_{i=1}^{n}(X_i - Y_I)^2}$

Where X_i and Y_i are existing points.

b) The formula of Manhattan distance is - $d = \sum_{i=1}^{n}(|A_i - B_i|)$

where A_i and B_i are the existing points.

c) The formula of Minkowski distance is – $d = \sqrt{\sum_{i=1}^{n}(X_i - Y_I)}^{1/p}$

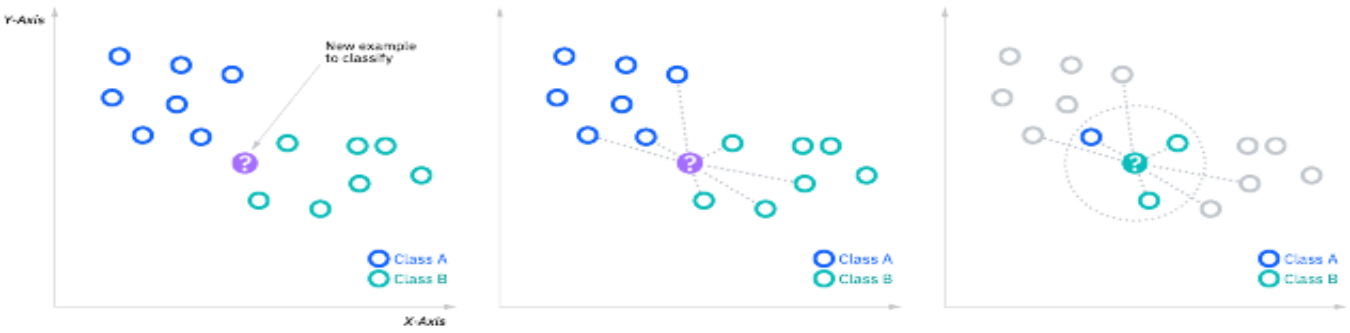


Figure 2: Kth near neighbour's work flow

3) SVM-

SVM is known as support vector machine it is a supervised machine learning model. Here, try to find a hyper plane that separate the two classes and also find out the best hyperplane. SVM is basically finding the maximum margin between the hyperplanes that means the maximum distance between the two classes. Two types of SVM models are used.

a) Linear SVM-

Separate the data points by a single straight line is called Linear SVM.

b) Nonlinear SVM-

when the data points cannot be separated by a line then, kernel trick are effectively classifying the data points. In real world application data points are often not linearly separable, making kernel trick essential. Different types of kernel trick are utilized to achieved better classification result.

Gaussian Kernel, Gaussian Kernel Radial Basis Function, Sigmoid Kernel & polynomial Kernel.

Mostly Gaussian Kernel RBF function is used.

- **Margin-**

  In classification task the margin refers to the distance between the decision boundaries and the closest data point from each class. In SVM large margin is called good margin. The two types of margins are hard margin and soft margin.

- **Support Vector-**

  The points which are closest to the hyper plane are called support vector.

  Two types of hyperplanes are there one is called positive hyper plane and other is called negative hyperplane.

- **Hard-Margin-**

  If the training data is linearly separable in this case classifier aims to find the maximum margin hyperplane that strictly separate the classes without allowing miss classification. However, it is very sensitive for the outliers. The equation of the hyperplanes are-

  $W^T x - b = 1$ & $W^T x - b = -1$[any points that below the boundary is one class and other is above boundary, with label -1, 1]

  Geometrically, the distance between these two hyperplanes is $2/\|w\|$, so to maximize the distance between these two hyperplanes need to minimize the value of $\|w\|$. The distance is computed using the distance from a point to a plane equation.

  $W^T x - b >= 1$, if $y_i = 1$
  $W^T x - b <= -1$, if $y_i = -1$

This can be written as $y_i(W^T x - b) >= 1$ for all $1 <= i <= n$

- **Soft-Margin-**

  To extend SVM to cases in which the data are not linearly separable.

  $\max (0, 1-y_i (W^T x_i - b))$

  Note that y_i is the i-th target and $(W^T x_i - b)$ is the i-th output.

  The goal of the optimization is to minimize-

  $\|w\|^2 + C [1/n \sum_{i=1}^{n} \max(0, 1-y_i (W^T x_i - b))]$

  Where C>0 determines the trade off between increasing the margin size and ensuring that the x_i lie on the correct side of the margin.

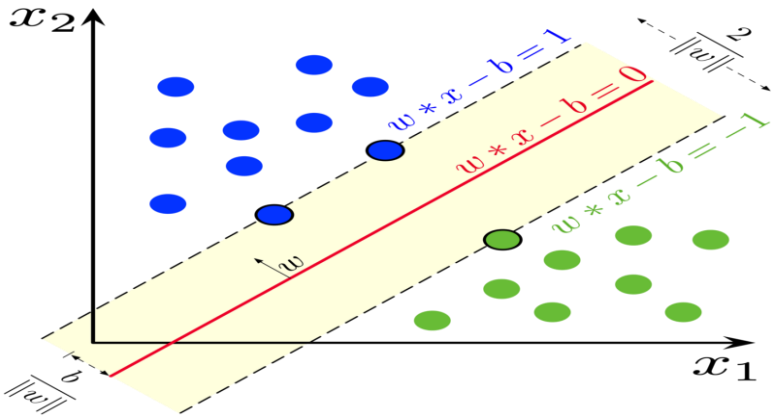


Figure 3: Support Vector Machine work flow

4) Ada Boost-

Ada boost algorithm is known for adaptive boosting, is a boosting technique used as an Ensemble learning technique. It is a supervised machine learning algorithm that is used to classify data by combining multiple weak learners into a strong learner. In Ada boost Decision stumps used as weak learners.

a) Weak learner-

When the model accuracy is over 50% then it is called weak learner. Although individually weak , they contribute to a strong learner when combined through boosting.

b) Decision Stumps-

The decision tree which's max depth is 1is called decision stumps. Decision stumps are ideal as weak learners in Ada boost due to low complexity and rapid training.

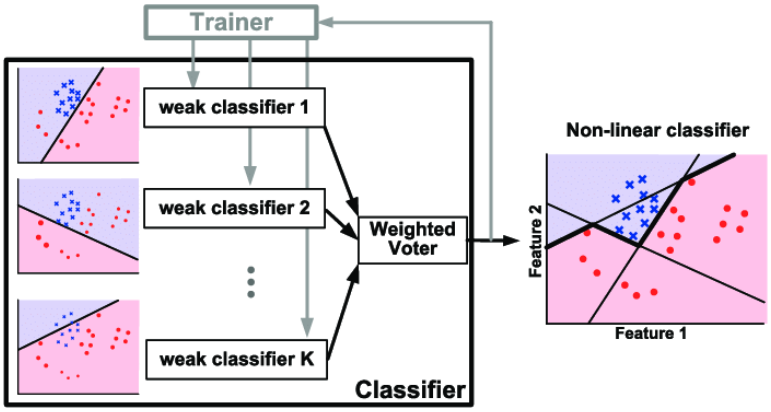


Figure 4: Ada Boost work flow

5) Cat Boost-

Cat boost is a gradient boosting frame work that is specially designed for categorical feature support and its known for the good performances. It was developed by Yandex a Russian IT company. The name “Cat Boost” came from categorical boosting.

Cat boost is very useful for categorical features. The most widely used technique which is usually applied to low cardinality categorical features is one-hot encoding: In this process the original feature is removed and the new binary number is added for each category. In the training phase One-hot encoding has been done to change categorical variable to binary, latter can be implemented more efficiently in terms of training time and is implemented in Cat Boost. Another way to deal with categorical features is to compute some statistics using the label values.

Two main ideas make unique cat boost-

a) **Order boosting-**

Traditional gradient boosting faces a “prediction shift” problem where the model is overly trained on the training data and thus fails to generalize. This is due to “target leakage”, where information from the training example influence the model too much.

Cat boost solve the order boosting instead of learning of entire dataset at each step, cat boost build each iteration’s model by using only data up to the point, as though the data is arriving sequentially.

b) Handling Categorical Features-

Cat boost converts categorical data into meaningful numbers in a way that reduce the noise.

This keeps the model from getting biased by rare categories and improve its ability.

6) XG Boost –

XG Boost known for Extreme Gradient Boosting, where the term Gradient boosting originates from Greedy Function Approximation. XG Boost is a generalized gradient boosting implementation that includes a regularization term, used to combat overfitting as well as support for arbitrary differentiable loss function. It utilizes the decision tree as a base learner and employs regularization techniques to enhance the model generalization. It is known for computational efficiency, feature importance analysis and handling of missing values. This model is widely used for regression, classification and ranking. Ensemble learning offers systematic solution to combine to combine the predictive power of multiple learners.

XG Boost introduce a regularization term to prevent overfitting which helps in producing simpler model that generalize better. This model implements a "weighted quantile sketch" for efficient split finding in large dataset and use sparsity- aware algorithms to manage missing values and optimize for sparse data.

The objective function of XG Boost model is-

$L(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k)$

Here l is a differentiable convex loss function that measures the difference the prediction between $\hat{y}_i$ and the target y_i. The second term $\Omega(f_k)$ describe the complexity of the tree f_k and it define as-

Where $\Omega(f_k) = \gamma T + 1/2 * \lambda|w|^2$

The additional regularization term helps to smooth the final learnt weights to avoid the overfitting. γT provides a constant penalty for each additional tree leaf and $\lambda|w|^2$ penalizes extreme weights.

7) HOG feature extraction-

a) First, the input image scaled into 64 X 128 pixels to ensure that consistent size of image is used in the feature extraction. The image is then converted into gray scale.

b) Then calculate the gradients for each pixel.

$dx = I(x+1, y) - I(x, y)$

$dy = I(x, y+1) - I(x, y)$

where dx & dy are the horizontal and vertical gradient respectively, I (x, y) is the pixel value of (x, y) position and calculate the gradient orientation θ

$\theta(x, y) = \tan^{-1}(dy/dx)$

c) Then the resulting gradient image is divided into cells of 8 X 8 pixels. A sliding window with the size of 16 X 16 pixels is then slide through the cells, with each step covering 4 neighboring cells. Each group of 4 neighboring cells creat block. The neighboring blocks are overlapping with each other. Through this process, a total of 105 blocks are formed on the 64 X 128 pixel's image. The division of image into blocks is similar to the original HOG.

d) For each block, a histogram for gradient orientation is constructed. This is done by voting of the orientation angles of each pixel into a predefined number of histogram bins. Using the higher number of bins will extract more detailed orientation information from the image but it will generate higher number of features. In order to reduce feature size and yet to retain the important details in the feature different number of histogram bins is used for different region in the image. Higher number of histogram bins is used for different regions in the image. Higher number of histogram bins is used extract features for the region s which may belong to part of a human figure while lower number of bins is used for the rest of the region. Higher number of histogram bins is used to extract feature for the shaded blocks while lower number of histogram bins is used for the rest of blocks. The optimal values for high and low number of bins to be used are determined empirically and experiment results are determined empirically.

e) L2 normalization scheme is used to normalize the histogram values to make the features more invariant to change in illumination. Here L2 norm is the best normalization scheme for HOG feature extraction.

$$v_n = v / \sqrt{\|v\|_2^2 + \epsilon^2}$$

where v is the un-normalized feature vector, v_n is the normalized feature vector $\|v\|_2$ is the L2 norm value and ϵ is the small value of constant to avoid devision by zero.

The value of L2-norm is calculated by the following equation-

$\|v\|_2=\sqrt{\sum_{i=1}^{n} v_i^2}$

Multiple block normalization is performed by grouping several neighbouring blocks according to their number of orientation bins used in the histogram generation.

8)LBP feature Extraction-

Local binary pattern method has been used in various applications. The LBP has been used various applications. LBP is an operator for texture description that based on the signs of differences between neighbor pixels and central pixels. For each pixel value in the image, a binary code is obtained by thresholding its neighborhood with the value of the center pixel. This binary code can be considered as a binary pattern. The neighbor pixel becomes 1 if the pixel value is greater than or equal to the threshold value, and it becomes 0 if the pixel value is less than threshold.

$$S(x)=\begin{cases} 1 \text{ if } x \geq 0 \\ -1 \text{ if } x < 0 \end{cases}$$

Next the histogram will be constructed to determine the frequency values of binary pattern represents possibility of binary pattern found in the image. The number of histogram bins depends on the number of involved pixels in LBP calculation. If LBP uses 8 pixels, the number of histograms will be 2^8 which is equal to 256. The basic version of LBP operator uses the center pixel value as threshold to the 3x3 neighbor pixels. Threshold operation will create a binary pattern representing texture characteristic. The equation basic of LBP can be given below-

$LBP\ (X_i, Y_i) =\sum_{p=0}^{p-1} S(X_i - Y_i)2^p$

Where Y is the intensity value of the central pixel & X is the intensity of the neighboring pixel

9) Performance Evaluation-

- **Confusion Matrix-**

  A confusion matrix helps us gain insight into how correct our predictions were and how they help us the actual values.

	Actual Value	
Predicted value	TP	FP
	FN	TN

Where TP=True positive, FP=False Positive, FN=False Negative, TN=True Negative.

- **Precision-**

It is a parameter that is used to identify the positive predictions made by the model. It is formulated as true positive devided by complete true positive prediction.

Precision=TP/TP+FP

TP=True Positive

FP=False Positive

- **Recall-**

It is a performance measure of a model which is projected as the ratio of the total number of positive values which are predicted correctly divided by the total positive samples.

Recall=TP/(TP+FN)

TP=True Positive

FN=False Negative

- **F1 Score-**

The harmonic mean of precision & recall is called F1 score.

F1 score= 2(Precision * Recall/ (Precision + Recall)

- **Accuracy-**

This metrics actually measured the percentage of correct prediction to the total number of accurate predictions.

Accuracy= (TP+TN)/(TP+TN+FP+FN)

- **ROC-AUC-**

Receiver Operating Characteristic (ROC) curve and Area Under the curve (AUC) are widely used metrics evaluating the performance of binary or multi class classification models. ROC curve is a plot of the true positive rate (TPR) against the False Positive rate (FPR) for different threshold values. TPR is the ratio of correctly predicted positive observation to the all observation in actual positive class. FPR is the ratio of incorrectly predicted negative observation to the all observations in actual negative class. The ROC curve helps to visualize the trade off between sensitivity & specificity.

AUC provides an aggregate measure of the performance across all possible classification threshold. AUC ranges from 0 to 1, where higher AUC indicates better model performance. An AUC is 0.5 means no discrimination, while an AUC is 1 means perfect discrimination.

10) Multilayer perceptron-

The multilayer perceptron model is the most known and most frequently used type of neural network. On most occasions, the signals are transmitted with in the network in one direction: From input to output. There is no loop. The out put of each neuron does not effect the neuron itself. Layers which are directly not connected to the environment are called hidden layer. There are also feed-back networks are there which can transmit impulses in both directions, due to the reaction connections in the network. These types of network are very powerful and can be extremely complicated. They are dynamic, changing their condition all the time, until the network reaches the equilibrium state and the search for a new balance occurs with each input change. A perceptron with a single layer and one input generates decision regions under the form of semi planes. By adding another layer, each neuron acts as a standard perceptron for the outputs of neurons in the anterior layer. Thus, the output of the network can be estimate as convex decision regions. A three layer perceptron can generate arbitrary decisions areas. Regarding the activation function of neurons, it was found that multilayer networks do not provide an increase in computing power compared to networks with a single layer, if the activation function are linear. The power of Multilayer perceptron comes precisely from non-linear activation function. Almost any non-linear function can be used for this purpose, except for polynomial function.

- **Input Layer –**

It is initial or starting layer of the multiple perceptron. It takes input from training data set and forewords it to the hidden layer. There are n inputs nodes in the hidden layer. The number of input nodes depend on the number of dataset features. Each input vector variable is distributed to each of the nodes of the hidden layer.

- **Hidden Layer-**

It is the heart of all artificial neural networks. The layer comprises all computations of the neural network. The edge of the hidden layer have weights multiplied by the node values. This layer uses activation function. There can be one or two hidden layers in the model.

- **Output Layer-**

This layer gives the estimated output of the Neural Network. The number of nodes in the output layer depends on the type of problems. For a single targeted use one node. For N classification problem, ANN uses N nodes in the output layers.

- **Activation Function-**

In MLP's some neurons use a nonlinear activation function that was developed to model the frequency of action potentials, of biological neurons. The two historically common activation functions are both sigmoid, and are describe by-

$Y(v_i)=\tanh(v_i)$ & $Y(v_i)=(1+e^{-v_i})^{-1}$

The first function is a tangent function that ranges from -1 to 1.

While the other is a logistic function that's ranges from 0 to 1.

Here Y_i is the output of ith node and v_i is the weighted sum of the input connection. In recent developments rectified linear unit is more frequently used as one of the possible ways to over come the numerical problems related to sigmoid.

- **Layers-**

The MLP consists three or more layers of nonlinearly activation nodes. Since MLPs are fully connected, each node in one layer connects with a certain weight w_{ij} to every node in the following layer.

The first type of layer is the Dense layer, also called fully connected layer and is used for abstract representations of input data. In this layer neuron connect to every neuron in the preceding layer. In MLP these layers are stacked together.

The convolution layer is typically used for image analysis tasks. In this layer, the network detects edges, textures, patterns. The outputs from this layer are then fed into a fully connected layer for further processing.

The pooling layer is used to reduce the size of data points. The recurrent layer is used for text processing with a memory function. Similar to the convolutional layer, the output for recurrent layers are usually fed into a fully connected layer for further processing.

The normalization layer adjusts the output data from previous layers to achieve a regular distribution. This results in improved scalability and model training.

.

11) VGG 16-

The network use 3*3 convolution filter, the smallest filter size that capture special hierarchies with in each layer. This choice balances depth and computation efficiency minimizing the

number of parameters compared to large filter while stacking layers increase non-linearity enhancing pattern recognition ability.

Max-pooling with 2*2 filter and a stride of 2 reduce the spatial dimensions after certain convolutional layer and three fully connected layers follow the convolutional layers.

This vgg 16 is a convolutional neural network with up to 19 layers using small 3*3 convolutional filter. This architecture achieved state of the art result on the image net datasets. For training vgg16 used a fixed 224*224 input image size, batch normalization and dropout for regularization with the final soft max layer for classification. Multi-scale testing, data augmentation and network fusion technique further enhance its performance. The network generalizes well to other data sets showing significant performance improving in various image classification task and influencing advancement in application such as object detaction, feature extraction and Image segmentation.

❖ Experimental Results –

In this paper ISIC 2017 dataset was used where 2357 images were found. melanoma, pigmented benign keratosis, vascular lesion, basal cell carcinom, squamous cell carcinoma, nevus, dermatofibroma, actinic keratosis, seborrheic keratosis these 9 classes were found.

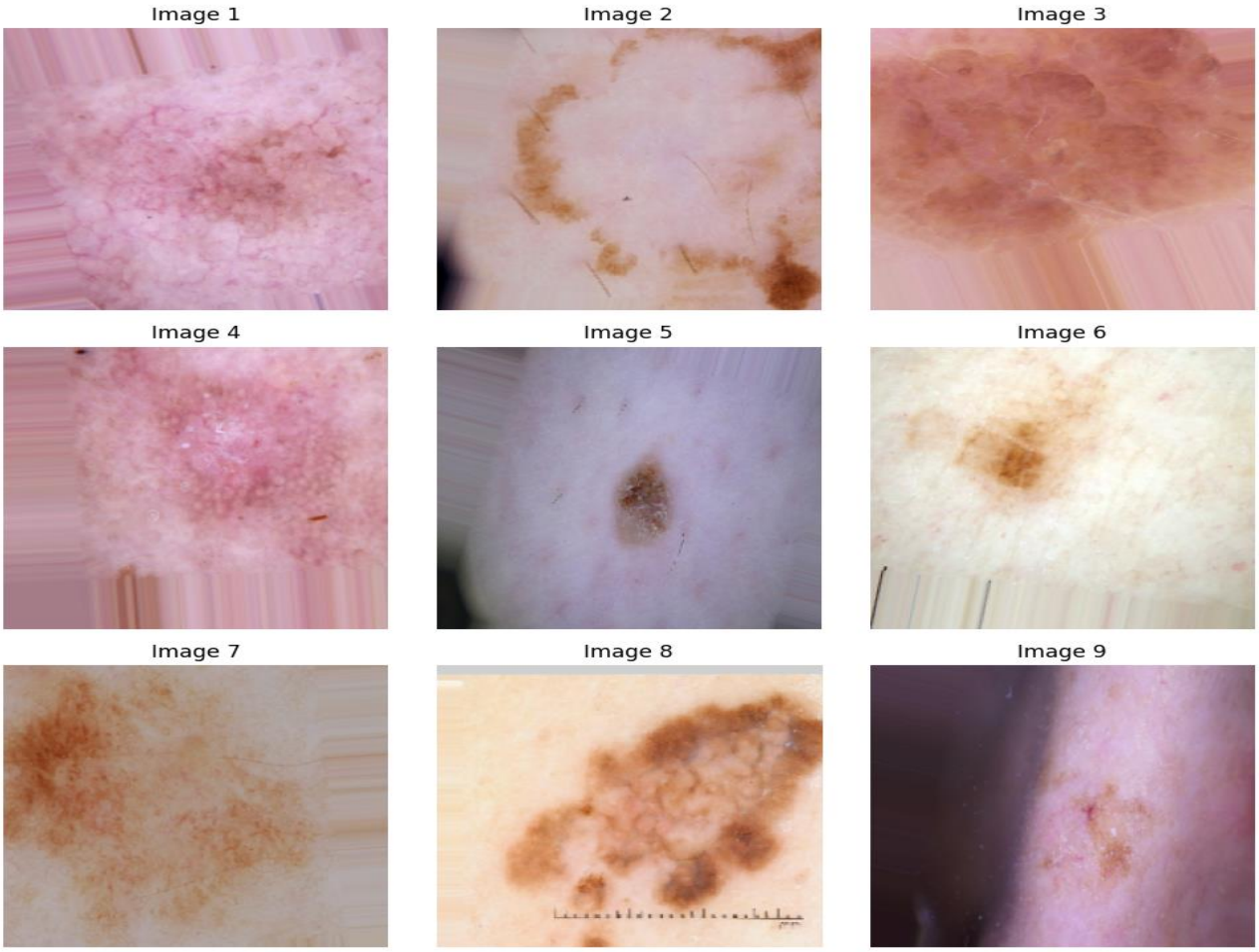


Figure5: resized images

These are the 9 classes that used in this paper. those pictures were after resized, noise detection from the datasets.

In this paper for this project used various kind of machine learning and deep learning models for multiclass skin cancer classification using ISIC 2017 dataset. The model tested including Decision tree, KNN, Support vector machine, Ada Boost, XG Boost, Cat Boost, and a Multi-layer Perceptron. This paper aimed best performance model based on Accuracy, precision, recall score.

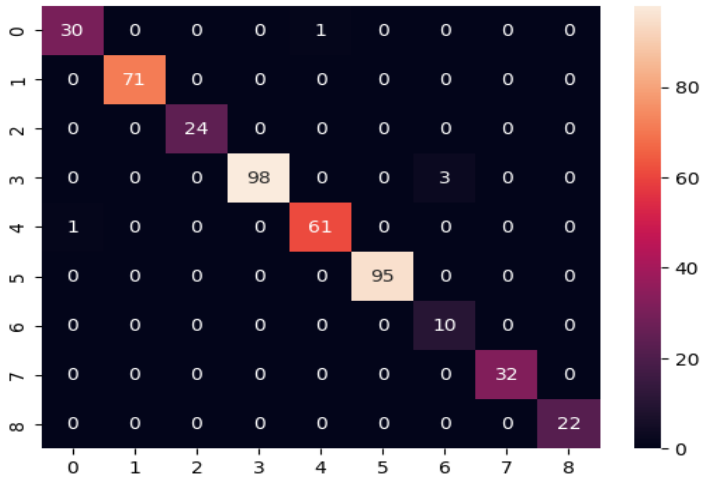
Model Name	Train Accuracy	Test Accuracy	Precision	Recall	F1-score
Decision Tree	92.23%	95.08%	0.95	0.92	0.93
KNN	57.73%	55.13%	0.47	0.49	0.45
Support Vector Machine	75.48%	67.63%	0.65	0.47	0.46
Ada Boost	38.52%	30.35%	0.17	0.32	0.13
XG Boost	95.47%	98.89%	0.97	0.99	0.98
Cat Boost	77.83%	95.75%	0.91	0.97	0.93
Multi-Layer Perceptron	96.72%	77%	0.78	0.77	0.77

- **Best performance model –**

The XG Boost model demonstrated the best performance across most metrics, achieving a train accuracy of 95.47% and a test accuracy of 98.89%. It also recorded highest precision (0.97) and recall (0.99) among all the models. These results indicate that Xg Boost effectively generalized to the test data with minimal overfitting, making the optimal choice for this classification task.

- **Other notable model-**

Decision Tree and Cat Boost models also showed strong performance with high test achieving accuracy of 95.08% and 95.75% respectively.



Confusion matrix-1: XG Boost Model

From the above confusion matrix, it highlighted that the distribution of across classes for the best performing model, XG Boost. The matrix shows the model is mostly correctly classified majority of classes.

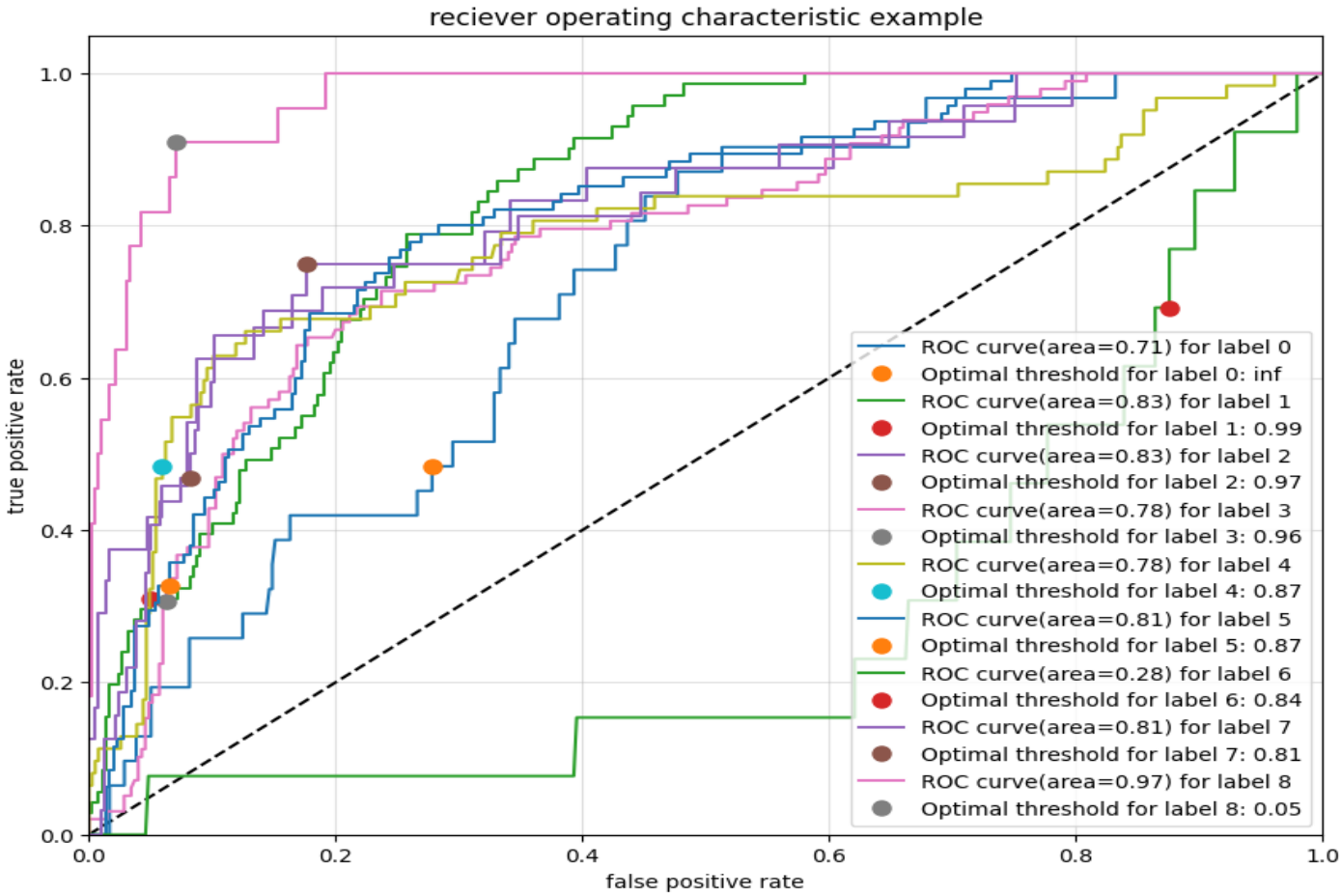


Figure6: AUC-ROC Plot

From the above ROC-AUC plot it shows that the AUC values from each label indicate that model's ability to distinguish between classes. Higher AUC values suggest better discrimination between melanoma and non-melanoma case for each specific label.

For instance, values around 0.83 to 0.97 indicate relatively strong performance.

The threshold values shown in the plot represent the points where the model achieves a balance sensitivity and specificity for each label.

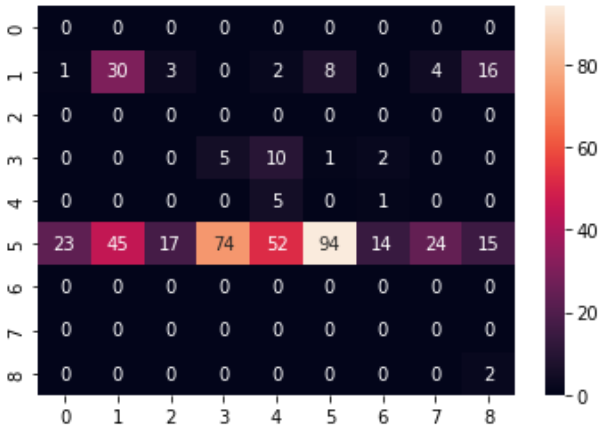
Lower threshold (close to 0) may yield higher sensitivity, meaning more melanoma cases are correctly identified. Higher threshold leads to higher specificity, that means fewer non-melanoma cases are incorrectly labeled as melanoma.

For melanoma detection, higher sensitivity by selecting lower thresholds where the model is more likely to classify cases as positive. This approach ensure that the model catches the

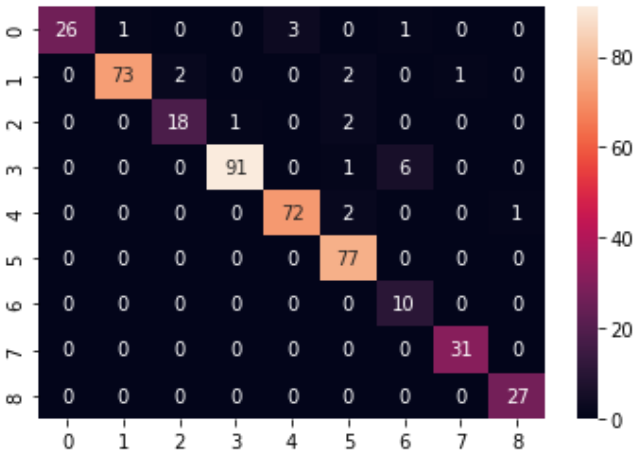
majority of melanoma cases. This sensitivity – focused threshold selection is very crucial as early detection of melanoma significantly improved treatment outcome.

Here, need of balance sensitivity with specificity to avoid excessive false positive, which can lead to unnecessary procedures and patient's anxiety. By adjusting threshold based on the ROC insights, aiming to create a melanoma detection model that aliens in clinical

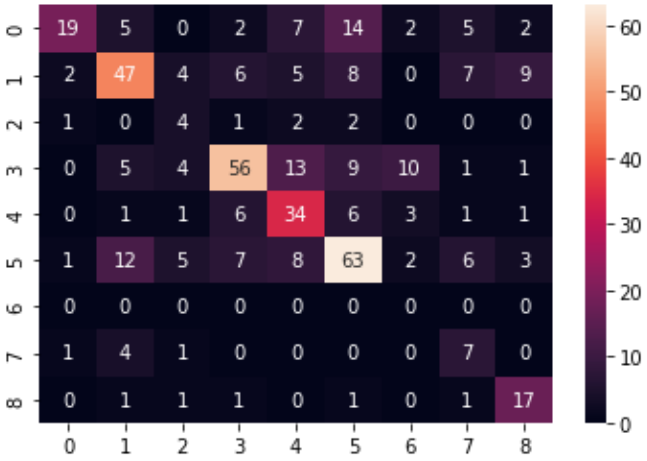
Priorities.



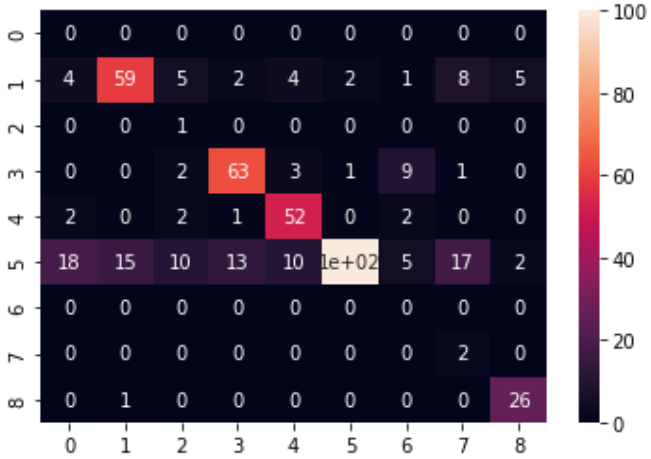
Confusion matrix2: Ada Boost



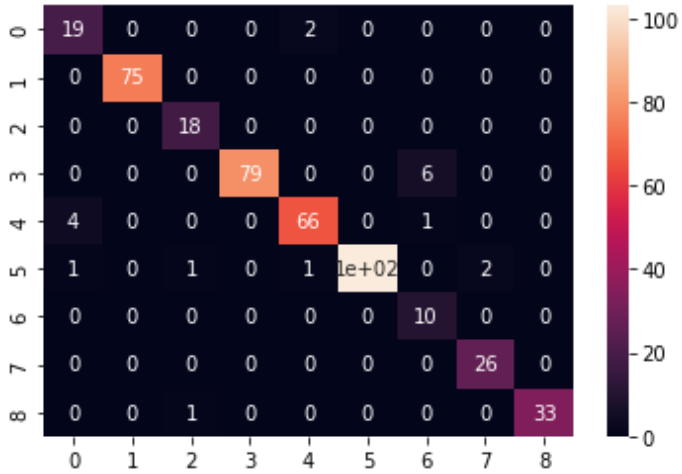
Confusion matrix 3: Decision Tree



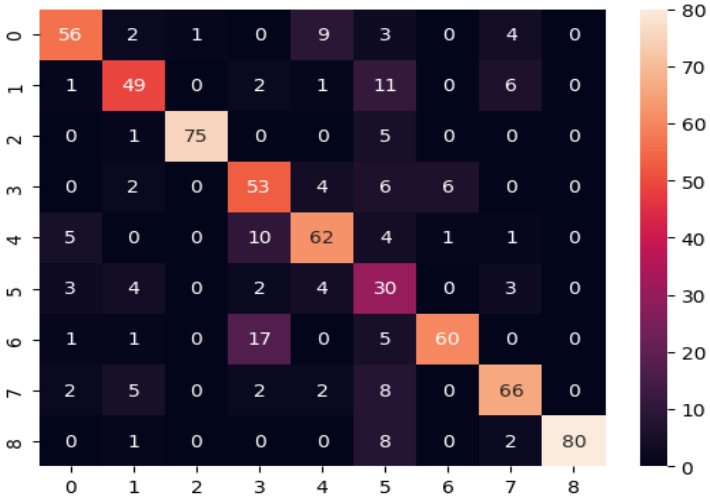
Confusion matrix 4: KNN



Confusion matrix 5: Support Vector Machine



Confusion matrix 6: Cat Boost



Confusion matrix 7: Multi-layer perceptron Model

❖ Conclusion-

Melanoma cancer is most dangerous from the other skin cancer with a high mortality rate. Due to its rapid spared in globally, the early detection is very important to reduce the mortality rate. Therefore, there is an urgent need for efficient and accessible diagnostic tools that can detect melanoma as early as possible.

To addressed this need, we develop a multi class classification model based on XG Boost model designed to accurately identify melanoma and other type of skin cancer. After pre-processing, feature extraction and label encoding this model achieved impressive 98.89% accuracy out performing other models in its ability to classify skin cancer. The high accuracy suggests that XG Boost model distinguish between melanoma and other skin cancer which is crucial for early and accurate diagnosis.

In future Implementing this model with computer aided diagnostic (CAD) system we can provide dermatologist with a powerful tool to enhance their decision-making capabilities.
The deployment of such system could have profound implications for reducing melanoma related mortality rates. By providing with dermatologist with accurate insight and potential cases of melanoma, this model could minimize human error and also minimize delayed diagnosis. Further-more this approach aligns with the goal of enhancing accessibility of healthcare as automated system can operate in areas with limited access to specialized medical professionals, ensuring that patient is remote or underserved region also benefits of early detection of melanoma.

REFERENCES-

1. Rashid, J., Ishfaq, M., Ali, G., Saeed, M. R., Hussain, M., Alkhalifah, T., ... & Samand, N. (2022). Skin cancer disease detection using transfer learning technique. *Applied Sciences*, *12*(11), 5714.

2. Gouda, W., Sama, N. U., Al-Waakid, G., Humayun, M., & Jhanjhi, N. Z. (2022, June). Detection of skin cancer based on skin lesion images using deep learning. In *Healthcare* (Vol. 10, No. 7, p. 1183). MDPI.

3. Monika, M. K., Vignesh, N. A., Kumari, C. U., Kumar, M. N. V. S. S., & Lydia, E. L. (2020). Skin cancer detection and classification using machine learning. *Materials Today: Proceedings*, *33*, 4266-4270.

4. Ansari, U. B., & Sarode, T. (2017). Skin cancer detection using image processing. *Int. Res. J. Eng. Technol*, *4*(4), 2875-2881

5. Allugunti, V. R. (2022). A machine learning model for skin disease classification using convolution neural network. *International Journal of Computing, Programming and Database Management*, *3*(1), 141-147.

6. Kousis, I., Perikos, I., Hatzilygeroudis, I., & Virvou, M. (2022). Deep learning methods for accurate skin cancer recognition and mobile application. *Electronics*, *11*(9), 1294.

7. Ozkan, I. A., & Koklu, M. (2017). Skin lesion classification using machine learning algorithms. *International Journal of Intelligent Systems and Applications in Engineering*, *5*(4), 285-289

8. Ameri, A. (2020). A deep learning approach to skin cancer detection in dermoscopy images. *Journal of biomedical physics & engineering*, *10*(6), 801.

9. Rahman, M. M., Nasir, M. K., Nur, A., Khan, S. I., Band, S., Dehzangi, I., ... & Rokny, H. A. (2022). Hybrid feature fusion and machine learning approaches for melanoma skin cancer detection.

10. Alfi, I. A., Rahman, M. M., Shorfuzzaman, M., & Nazir, A. (2022). A non-invasive interpretable diagnosis of melanoma skin cancer using deep learning and ensemble stacking of machine learning models. *Diagnostics*, *12*(3), 726.

11. Gomathi, E., Jayasheela, M., Thamarai, M., & Geetha, M. (2023). Skin cancer detection using dual optimization based deep learning network. *Biomedical Signal Processing and Control*, *84*, 104968

12. Shorfuzzaman, M. (2022). An explainable stacked ensemble of deep learning models for improved melanoma skin cancer detection. *Multimedia Systems*, *28*(4), 1309-1323.

13. Ali, M. S., Miah, M. S., Haque, J., Rahman, M. M., & Islam, M. K. (2021). An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models. *Machine Learning with Applications*, *5*, 100036.

14. ijwil, M. M. (2021). Skin cancer disease images classification using deep learning solutions. *Multimedia Tools and Applications*, *80*(17), 26255-26271

15. Singh, P., Kumar, M., & Bhatia, A. (2022, May). A comparative analysis of deep learning algorithms for skin cancer detection. In *2022 6th international conference on intelligent computing and control systems (iciccs)* (pp. 1160-1166). IEEE.

16. Thurnhofer-Hemsi, K., & Domínguez, E. (2021). A convolutional neural network framework for accurate skin cancer detection. *Neural Processing Letters*, *53*(5), 3073-3093.

17. Emuoyibofarhe, J. O., Ajisafe, D., Babatunde, R. S., & Christoph, M. (2020). Early Skin Cancer Detection Using Deep Convolutional Neural Networks on Mobile Smartphone. *International Journal of Information Engineering & Electronic Business*, *12*(2).

18. Chaturvedi, S. S., Tembhurne, J. V., & Diwan, T. (2020). A multi-class skin Cancer classification using deep convolutional neural networks. *Multimedia Tools and Applications*, *79*(39), 28477-28498.

19. Jain, S., Singhania, U., Tripathy, B., Nasr, E. A., Aboudaif, M. K., & Kamrani, A. K. (2021). Deep learning-based transfer learning for classification of skin cancer. *Sensors*, *21*(23), 8142.

20. Hasan, M. R., Fatemi, M. I., Monirujjaman Khan, M., Kaur, M., & Zaguia, A. (2021). Comparative analysis of skin cancer (benign vs. malignant) detection using convolutional neural networks. *Journal of Healthcare Engineering*, *2021*(1), 5895156.

21. Bassel, A., Abdulkareem, A. B., Alyasseri, Z. A. A., Sani, N. S., & Mohammed, H. J. (2022). Automatic malignant and benign skin cancer classification using a hybrid deep learning approach. *Diagnostics*, *12*(10), 2472.

22. Gururaj, H. L., Manju, N., Nagarjun, A., Aradhya, V. M., & Flammini, F. (2023). DeepSkin: a deep learning approach for skin cancer classification. *IEEE Access*, *11*, 50205-50214.

23. Fraiwan, M., & Faouri, E. (2022). On the automatic detection and classification of skin cancer using deep transfer learning. *Sensors*, *22*(13), 4963.

24. Aljohani, K., & Turki, T. (2022). Automatic classification of melanoma skin cancer with deep convolutional neural networks. *Ai*, *3*(2), 512-525.

25. Garg, R., Maheshwari, S., & Shukla, A. (2021). Decision support system for detection and classification of skin cancer using CNN. In *Innovations in computational intelligence and computer vision: proceedings of ICICV 2020* (pp. 578-586). Springer Singapore

26. Babu, G. N. K., & Peter, V. J. (2021). Skin cancer detection using support vector machine with histogram of oriented gradients features. *ICTACT Journal on Soft Computing*, *11*(02), 2301-2305

27. Kausar, N., Hameed, A., Sattar, M., Ashraf, R., Imran, A. S., Abidin, M. Z. U., & Ali, A. (2021). Multiclass skin cancer classification using ensemble of fine-tuned deep learning models. *Applied Sciences*, *11*(22), 10593.

28. Suleman, M., Ullah, F., Aldehim, G., Shah, D., Abrar, M., Irshad, A., & Ayouni, S. (2023). Smart MobiNet: A Deep Learning Approach for Accurate Skin Cancer Diagnosis. *CMC- COMPUTERS MATERIALS & CONTINUA*, *77*(3), 3533-3549

29. Islam, M. K., Ali, M. S., Ali, M. M., Haque, M. F., Das, A. A., Hossain, M. M., ... & Rahman, M. A. (2021, April). Melanoma skin lesions classification using deep convolutional neural network with transfer learning. In *2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA)* (pp. 48-53). IEEE.

30. Naeem, A., Anees, T., Fiza, M., Naqvi, R. A., & Lee, S. W. (2022). SCDNet: a deep learning-based framework for the multiclassification of skin cancer using dermoscopy images. *Sensors*, *22*(15), 5652