

# EVALUATION OF MACHINE LEARNING AND DEEP LEARNING APPROACHES TO CLASSIFY BREAST CANCER USING THERMOGRAPHY

**P.KANIMOZHI<sup>1</sup>, S.SATHIYA<sup>2</sup>, M.BALASUBRAMANIAN<sup>3</sup>, P.SIVAGURU<sup>4</sup>, P.SIVARAJ<sup>5</sup>**

1PhD scholar in CIS, Annamalai University, Tamilnadu, India, khani.pappi@gmail.com

2 Assistant Professor in CSE, Annamalai University, Tamilnadu, India, sathiya.sep05@gmail.com

3 Associate Professor in CSE, Annamalai University, Tamilnadu, India, balu\_june1@gmail.com

4 Assistant Professor in civil, University college of Engineering, Pattukottai, rpsivagurucivil@gmail.com

5 Associate Professor in MFG, Annamalai University, Tamilnadu, India, cemajorsiva@gmail.com

## ABSTRACT

Emerging technologies undergo the challenge of diagnosing fatal diseases even before any symptoms could occur in human body. One such syndrome is breast cancer which is now common in developing countries like India and increasing rapidly among women. The only remedy is diagnosing the disease earlier using latest technology like thermography, one of the best methods for cancer screening test without any harmful effect. This paper concentrates on various sectors namely segmentation where two approaches are discussed, next is feature extraction method used to extract the vital features for classification and last is classification technique employed for both machine learning and deep learning methods. Of all the models discussed Residual Network model proves highest accuracy of 96% in classification. The models performance are related by training the thermal images and a comparative analysis is performed.

**KEYWORDS:** Thermography, Segmentation, Classification, feature extraction and Residual Network model.

## 1. Introduction

A heterogeneous disease which once developed in human body sustains and spreads to other parts finally leading to death is breast cancer. Breast cancer threatens mostly female patients and rarely male victims are also present. The cells in human body when grow abnormally even without any need, develop into a lump or mass called a tumor. Such tumors when advanced in the breast region are called breast cancer which spreads faster to other regions. Usually the breast is comprised of mammary glands, fat, blood vessels, nerves and lymphatics.[1]. Breast cancer first forms in the ducts and the lobules of the breast and spreads to neighbour tissues also by joining through blood or lymph systems. So the tumor or a mass is first detected near the nipple and the darker region of the breast called areola. The only method to fight against the disease is the early screening method to diagnose at the very primary stage. An accurate method for diagnosing of breast cancer before any symptom visible is the Thermography method. Two of the techniques are

implemented using thermal images to detect breast cancer, one is the machine learning method applied. Another new improved technique used is the deep learning method to compare the results for better accuracy.

## 2. Related work

This section deals with related work on diagnosis of breast cancer based on the diversity of various researches conducted so far. The following section deals with the achievement of various works in this field.

Although in prior, many methods were in use standard mammogram methods are considered in spite of their disadvantages. According to Saira Charan [2] Convolutional Neural network model is used in mammogram for breast cancer detection. Training and testing the data set with CNN layers are done to classify normal and abnormal images. In [3] a segmentation method based on deep learning

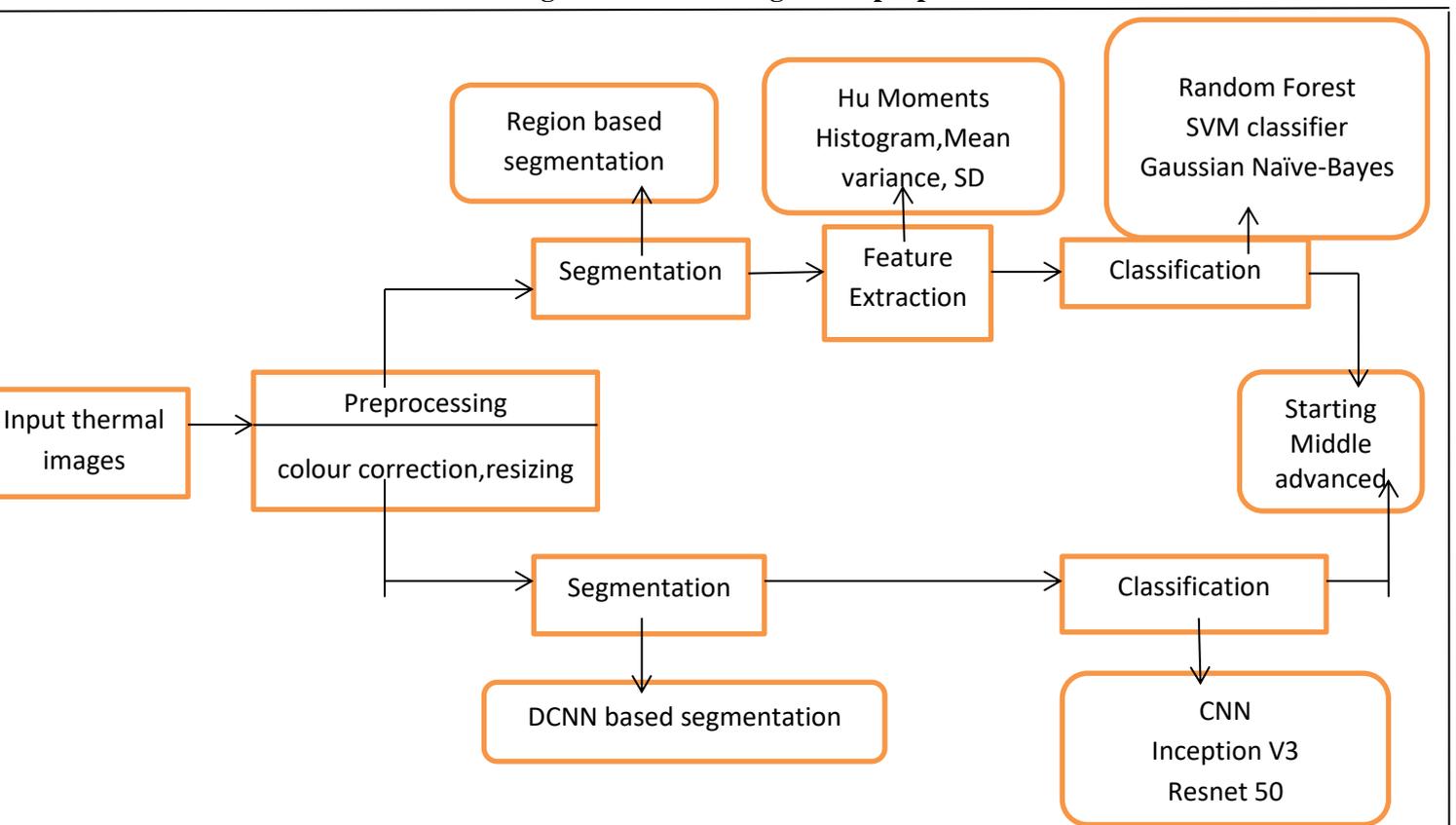
architecture for ultra sound imaging is used with BUS image segmentation and the data is trained with two-fold cross validation. By Dina et.al.[4] region based and threshold method are used in first phase while Alexnet with SVM are used for feature extraction and classification to classify normal and abnormal images of breast cancer. Sharkas, Al-Sharkawy & Ragab (2011) used the discrete wavelet transform (DWT), the contourlet transform, and the principal component analysis (PCA) methods for feature extraction. The system was able to detect and classify normal and abnormal tissues. The achieved rate was almost 98%. Dhungel, Carneiro & Bradley (2015) used the multi-scale belief network in detecting mass in mammograms. The sensitivity achieved was 85%–90% using the INbreast and DDSM-BCRP datasets, respectively.

### 3. Proposed work

The proposed work aims at providing the accurate method for diagnosing breast cancer using thermal images and comparing the results of machine learning methods with deep learning techniques. In both the methods input images are preprocessed and segmented to divide images as normal and abnormal classes. Features are extracted and preprocessed in machine learning method and given as input to classifiers to categorize abnormal images into three classes namely starting, middle and advanced based on stages. In deep learning hyper parameters are extracted and used for classification.

The block diagram of the proposed work is given in the figure 3.1 below.

**Figure 3.1. Block diagram of proposed work**



The proposed method contains the steps mentioned above in block diagram Fig 3.1. Finally the results obtained from both the methods after classification steps are compared to obtain better accuracy.

### 3.1. Image Acquisition

Thermal images are collected from various databases like DBT-TU-JU database, Ann Arbor Thermography, DMR-IR-database. Thermal images are produced by special Infra red cameras like FLIR T400 with temperature range of -20 to 1700° c used for image capturing in JPEG Format.

### 3.2. Image Preprocessing

The foremost step is the preprocessing of images which is not needed in high level for thermal images. To prepare the images for segmentation accurately the thermal images undergo two methods. First colour correction method is applied where the thermal images are converted into RGB colour model images. Next the thermal images are resized to a standard size of 128X128 which can be done using Matlab coding in the algorithm proposed. These preprocessing steps are implemented in both the segmentation methods.

### 3.3. Image Segmentation

Segmentation is the process of segmenting that is breaking the images into uniform regions based on some criteria to crop out the affected portion from the background of the image. In the proposed model there are two novel methods discussed in the following sections to segment the images to identify whether the images belong to normal or abnormal persons.

#### 3.3.1. Region based Segmentation Method

Thermal images are segmented using region based segmentation method with hierarchical codebook and

#### 3.3.3 Experimental Results for Region based Segmentation method

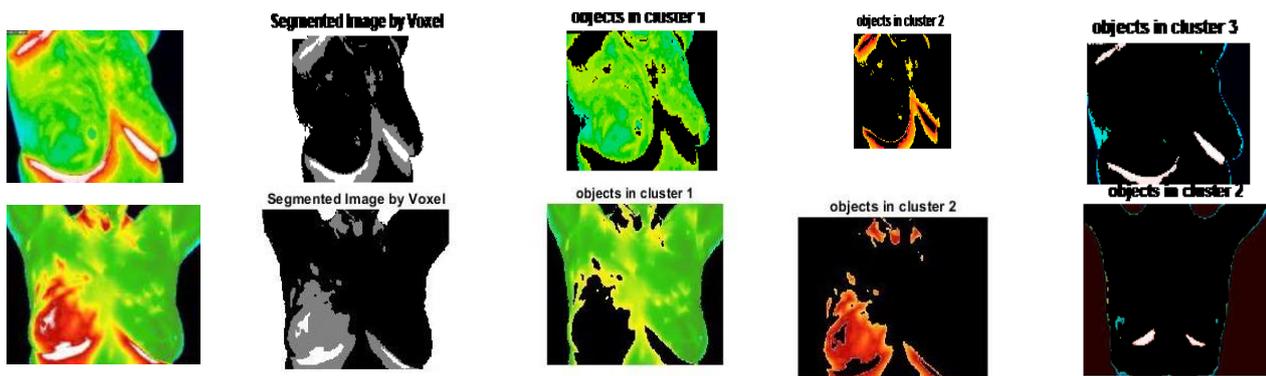
supervoxel as the criteria. For this method 3D thermal images are used for better accuracy and codevectors are used to represent the predominant intensity of the pixel values. This method is based on the fact that rise in temperature in the human body is indicated by a change in the intensity of the pixel values clearly indicated by the color differentiation like red stands for high intensity and light colours like blue stand for low intensity of the pixels. More details are in paper[5] discussed in detail.

#### 3.3.2. DCNN based Segmentation method

The second method applied for segmentation is Deep Convolutional Neural Network method used for segmentation of thermal images. The method is based on the modified U-Net architecture based on data augmentation and trains very few images with good accuracy.

The architecture of U-Net model is constructed with a contracting path to capture the details of the input image and a symmetric expanding path for locating the particular region of interest. The left part of the model is contracting path done by convolutional process and right side is expansive path done by reverse 2D convolutional layers. In the proposed model there are 5 convolutional layers with 1 max pool layer between 2 convolutional layers in the left part. There are 4 concatenate functions to get more precise prediction. The last step is to reshape the image which is a 1x1 convolution layer to produce the output of size 128X128X3. The model after trained with the thermal images produce a binary segmentation mask by giving the threshold function called binary inverter threshold of value 255 with the value 255 as maximum and 49 as minimum to retrieve the cancer affected region indicated by red colour pixels. The mask is added with the input image to produce the required segmented output image.

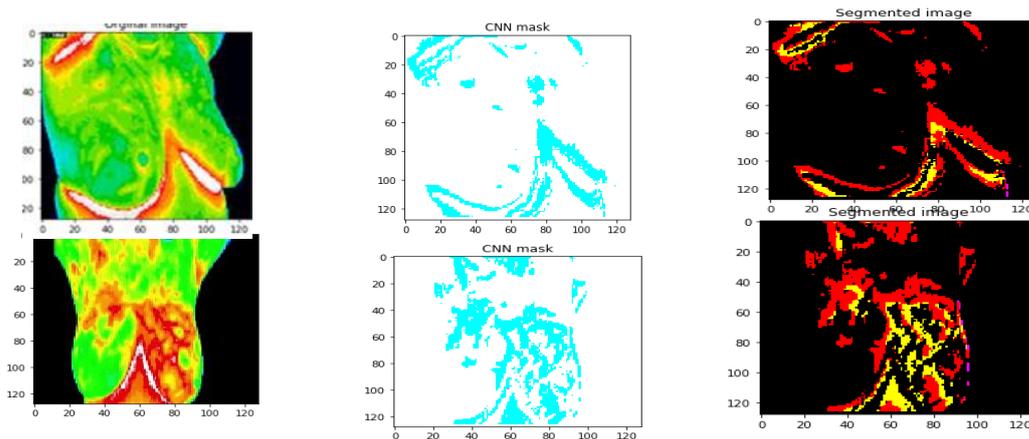
Figure 3.3.3.1 Segmented results of normal and abnormal images



First row gives normal images which is clearly shown in the figure and the second row gives the abnormal image with the cancer affected regions segmented using region based method.

3.3.4. Experimental Results for DCNN based segmentation method

Figure 3.3.3 Segmented results of normal and abnormal images



The first row shows the segmented part and the mask of normal image. Second row shows abnormal image with the mask and the output segmented image showing cancer affected areas using DCNN method.

3.4. Feature Extraction

The crucial stage is the feature extraction method where after segmentation the images are classified using the features that are extracted from the segmentation methods discussed above. Basically an image contains features like colour, texture, shape, position, dominant edges of image items and regions. Some of the methods discussed are Hu Moments, Color Histogram and first order statistical features like mean, variance and standard deviation.

3.4.1. Hu Moments

A moment is simply a number that gives the characteristics of an image usually represented in scalar quantities. Moments are used to differentiate the shape and their features like edges, lines, curves, regions and the color. In image processing, image moments are used to calculate the weighted average of the pixel intensities of the image to describe the objects after segmentation based on the properties of the image like total area or intensity, perimeter, the centroid and information about the orientation of the image. Hu moments belong to the image moments which uses raw moments, centroid localization, translation and scale invariance to solve the problem of shape identification. This method is very

suitable for 2D and recently 3D images also employ moment based method for object identification.[6].

Hu Moments ( invariants ) are a set of 7 numbers calculated using **central moments** that are invariant to image transformations. The first 6 moments have been proved to be invariant to translation, scale, and rotation. The feature vectors derived from this method are given as input to the classifier to categorize three classes of the segmented image. So Hu created some descriptors that are scale invariant from the geometric moments called raw moments. From raw moments, central moments are calculated and these values are used to calculate the 7 moments based on scale invariance. The seven Hu moments are the key point descriptors about the shape which contains the location and position given by translation invariance and scale invariance describe the pixel intensity values along with the size. The basic calculations are

1.Raw Moments- The total intensity or the weighted average of all the pixels in the image is calculated and their location in the image using geometric moments.. The first moment is calculated using the formula

$$M_{pq} = \sum_{x=0}^n \sum_{y=0}^n x^p y^q I(x, y)$$

Here  $I(x,y)$  gives basic 2d geometric moments of order  $(p+q)$  of the image

$x^p, y^q$  gives basics of the moments and  $p$  and  $q$  are weights of horizontal and vertical dimensions

The raw moments provide information about properties like area and size of the image

2.Centroid localization- The centroid coordinates are calculated to extract important information about the image. Here  $M_{00}$  accumulate non zero intensities and spatial information of the pixel.  $M_{10}$  and  $M_{01}$  are used to calculate the  $x$  and  $y$  dimensions and  $X$  and  $Y$  are the center coordinates, The formulas are

$$X = M_{10} / M_{00} \quad Y = M_{01} / M_{00}$$

With center coordinates the center point is taken

$M_{00}$  is total mass of the image,  $M_{01}$  and  $M_{10}$  are center mass of the image

3.Central Moments- Translation invariance where centroid rewrites raw moment to get translation invariant moments  $M_{pq}$  by calculating the value along

the centroid. The first two moments are used for computing the center points and central moments give relative spatial information of centroid taken. The formula is

$$M_{pq} = \sum_{x=0}^n \sum_{y=0}^n (x - x_c)^p (y - y_c)^q I(x, y)$$

This method is used to find the distance from center point to any point in the image.

4. Scale Invariance- The normalized central moments are the scale invariants used for the size changes that occurs in the image. Scaling can be uniform or non uniform and used to solve the problem arising due to different distances of the same object. Scale invariants are invariant to scale [7] and obtained using the formula

$$N_{pq} = M_{pq} / M_{00}^{1+p+q/2}$$

Translation and scale invariants are small and can be solved by themselves so the major problem is finding the rotational invariants and Hu defined the seven rotational invariants which are calculated from the central moments . They are

$$h_1 = N_{20} + N_{02}$$

$$h_2 = (N_{20} - N_{02})^2 + 4(N_{11})^2$$

$$h_3 = (N_{30} - 3N_{12})^2 + 3(N_{03} - 3N_{21})^2$$

$$h_4 = (N_{30} - N_{12})^2 + (N_{03} + N_{21})^2$$

$$h_5 = (N_{30} - N_{12})(N_{30} + N_{12})[(N_{30} + N_{12})^2 - 3(N_{03} + N_{21})^2]$$

$$+ (3N_{21} - N_{03})(N_{03} + N_{21})[3(N_{30} + 3N_{12})^2 - (N_{03} + N_{21})^2]$$

$$h_6 = (N_{20} - N_{02})[(N_{30} + N_{12})^2 - 7(N_{03} + N_{21})^2] + 4$$

$$N_{11}(N_{30} + N_{12})(N_{03} + N_{21})$$

$$h_7 = (3N_{21} - N_{03})(N_{30} + N_{12})[(N_{30} + N_{12})^2 - 3(N_{03} + N_{21})^2] +$$

$$(N_{30} - 3N_{12})(N_{03} + N_{21})[3(N_{30} + N_{12})^2 - (N_{03} + N_{21})^2]$$

All the above discussed raw and central moments are calculated for all the sample images and collected as feature vectors. The feature vectors are the input to classifier providing all the information of the contents of the images.

Totally 21 features are extracted using Hu moments and the table gives the representation of sample data set. Here Hu moments give the input as the segmented image and the output is the numbers zero to two indicating the three stages of cancer in the classification process.

**Table 3.4.1 Comparison of moments value for the segmented images**

S.NO	IMAGE STAGES	H[1]	H[2]	H[3]	H[4]	H[5]	H[6]	H[7]
1	STARTING IMAGE 1	1.01	2.68	3.11	1.73	3.83	2.51	1.23
2	STARTING IMAGE 2	1.01	2.68	3.12	1.73	3.84	2.50	1.20
3	MIDDLE IMAGE1	5.66	5.04	4.18	2.93	9.17	3.85	3.11
4	MIDDLE IMAGE2	5.67	5.04	4.17	2.91	9.17	3.85	3.11
5	ADVAMCE IMAGE1	8.00	3.35	1.07	4.82	7.60	8.13	7.92
6	ADVANCE IMAGE 2	8.00	3.36	1.07	4.80	7.60	8.13	7.90

### 3.4.2. Colour Histogram

The next method in feature extraction process is the Histogram method used for mining the pixel intensity values. Thermal image property is the raise in temperature due to absorption of heat which is indicated by the brightness of the pixel intensity values. These are the features extracted from the histogram to provide the input for classification method.

Histogram is a graph method consisting of the color properties of the pixels in an image with various intensity values that lies between 0 to 255 range. Histograms can be used in color images with red, green and blue channels either in a single or multiple histogram. This represents the pixel brightness values for every point in the image with pixel colors in x-axis and the number of pixel counts in the y-axis.[8]. For two dimensional image the features extracted are 256 and for thermal image that is three dimensional, 512 features are extracted from the images.

#### Histogram Algorithm

Histogram of an image is produced first by discretization of the colors in the image into a number of bins, and counting the number of image pixels in each bin. Histograms are constructed with bins or vertical bars of equal width made of data points which belong to separate class.

#### Algorithm

Histogram is calculated using the algorithm described below. The steps are

1. Load the thermal image
2. Split the image into its R, G and B planes using the function **cv::split**
3. Calculate the histogram of each one channel plane by calling the function **CV::calcHist**
4. Plot the three histograms in a window

The function cv:split is used to divide the input images into corresponding planes and cv: calchist is used for calculating the histogram of the given image. The histograms are normalised for better accuracy. The distribution of the red, green and blue colors derived from the histogram for the three stages are given in the table 3.4.1 below

**Table 3.4.1 Color distribution of thermal image in various stages**

IMAGES STAGE	RED PIXELS	GREEN PIXELS	BLUE PIXELS
STARTING	105-142	251-250	38-75
MIDDLE	120-205	125-194	41-90
ADVANCED	190-240	140-200	45-110

These are the feature vectors derived from the histogram from sample images given in the table used for classification.

**3.4.3. Statistical Features**

The first order statistical features called descriptive statistics like mean, variance and standard deviation are calculated as features from the entire image.

Mean is used for measuring the central tendency that is center point of the image and called as average value of the data in the image. It shows the contrast of one image with other image. So similarity of the images can be calculated. The formula is

$$\text{Mean} = x = 1/n \sum_{i=1}^N Xi$$

Where X represents the pixel values and n is the count of pixels

Variance is called measures of variation and used to calculate the distance from center point to particular point in average. Variance is calculated by finding the distance between data points and mean squaring and

summing and finally dividing with the number of points. The formula is

$$\text{Variance} = \sum_{i=1}^N \frac{(Xi-x)^2}{n}$$

Standard Deviation is the Square root of variance. It is used to find the dispersion within the local region that is edges of certain data points in the image. The formula is

$$\text{Standard.Deviation} = \sqrt{\sum_{i=1}^N \frac{(Xi-x)^2}{n}}$$

The values calculated for mean, variance and standard deviation in sample images are given in the form of table below. The values are used for comparison purpose for different stages of cancer.

**Table 3.4.2. Comparison of statistical features**

NO	IMAGES	MEAN	VARIANCE	Standard.Deviation
1	STARTING I	96.84	8271.90	90.95
2	STARTING II	91.61	7596.86	87.16
3	MIDDLE I	100.61	8981.35	94.77
4	MIDDLE II	102.12	9248.66	96.17
5	ADVANCED I	79.53	9508.20	97.51
6	ADVANCED II	76.84	10259.16	101.29

**3.5. Classification**

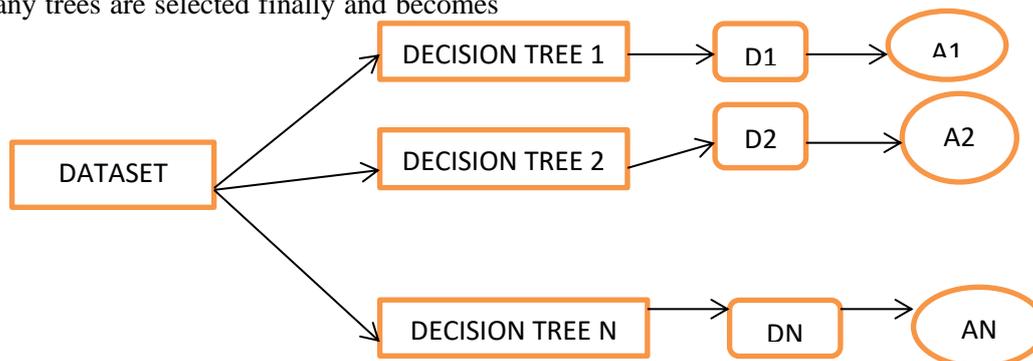
The success of best diagnosis system for identifying the disease lies in the classification method used. If the abnormality is classified precisely the classification methods used are up to the requirement. In this paper six types of classification methods are discussed for accuracy. First three methods namely Random Forest, Support Vector Machine and Gaussian Naïve Bayes method classify data set based on the features extracted with machine learning algorithms. Next CNN,

Inception V3 and Residual Network 50 models use deep learning methods for feature extraction and classification.

**3.5.1. Random Forest Method**

One of the cooperative techniques of combining multiple models in analogous and get the output is the Bagging method. The best example of Bagging method is the Random Forest classifier.

Random forest is a supervised learning algorithm where the model constructs different decision trees collectively called as forest. Bagging method ensures the accuracy when many models are trained with the features. Random forest method contains many decision trees in which internal node is the test on an attribute, each branch is the outcome of the test. The leaf node represents the final class. So the result of each decision tree produces a class prediction and the class predicted more from many trees are selected finally and becomes



**Figure 3.5.1.1. Sample random forest**

the model prediction. This step is called voting and majority voting class is predicted.

1. Bootstrap Aggregation- The method used to classify the data set into three classes. The individual tree is randomly sampled from dataset with replacement giving different trees. So multiple trees produces multiple results finally aggregated to find the majority.

## 2. How to construct random forest

1. Create a bootstrap data set by taking a small number of sample to build a tree from dataset.
2. Randomly select samples from the dataset
3. Create a DT using the bootstrapped dataset by using random columns at each step that is the subset of variables.
4. Select root node by taking important feature among the features extracted.
5. Different trees are constructed by taking features in different combination
6. Either among 2 , 1 feature is selected or among 3 , 1 feature is selected for root node.
7. Construct a tree with important feature as root node and check other features also one by one

Finally the sample should be classified by all the features selected to obtain which class the sample belongs. For prediction a new sample is tested with the trained tree. It is assigned the label of the training sample in the leaf node it ends up in. This algorithm provides many iterations with many decision trees and the average vote of all trees is reported as random forest prediction.

### 3.5.2. Support Vector Machine

The important terminologies of SVM are Support vector, Hyper planes selected, Marginal distance, Linear separable and Non-linear separable.

Support Vector machine is the supervised machine learning method used iteratively for classification problem. The model is the representation of various classes in the hyperplane in multidimensional space and divide the classes into various stages. The function used to classify the features in a 3-D is called as a plane similarly the function which classifies the point in higher dimension is called as a hyperplane which classifies the class based on the features extracted.

SVM builds and train a model using the data sets and classify the data points to find whether the sample belongs to three classes namely advanced, middle and starting stage. SVM works by mapping the data to a high dimensional feature space so the data points can be categorized by SVM classifier using kernel function. The kernel used in this method is Radial Basis Function kernel that expresses a measure of similarity between observations. The RBF kernel is defined as

$$K(a,b)= \exp (-\gamma(a-b)^2)$$

where  $\gamma$  is a parameter that sets the “spread” of the kernel. If the two observations are close together then,

$k(a - b)$  will be small. Then, so long as  $\gamma > 0$ , it follows that  $-\gamma (a - b)^2$  will be larger. Thus, closer vectors have a larger RBF kernel value than farther vectors.

SVM is suitable for binary classification but the data set should be classified into three classes so multiclass classification is done by extending the binary method using one vs rest method.

The three binary classification datasets are as follows:

- **Binary Classification Problem 1:** advanced stage vs. (middle and starting)
- **Binary Classification Problem 2:** middle vs. [advanced and starting]
- **Binary Classification Problem 3:** starting vs. [middle and advanced]

Binary classifiers train the model by treating the selected class as positive and all the other class data points as negative.

First the data set is divided into training and testing set. The model is trained by creating three different SVM models for three classes. The classifier is based on the features extracted and 3 different features are used for classification. Based on the pixel intensity values and shape features along with statistical features in first model, advanced stage data points are classified making the advanced class data points as true with all other data points as false. Similarly other two models are created for middle and starting classes. Apply testing data set to all the models and find the probability of data points that belong to any of the three classes based on the maximum probability value from the comparison of 3 models.

**3.5.3. Gaussian Naïve Bayes Method**

Gaussian Naive Bayes are a group of supervised machine learning classification algorithms based on the **Bayes theorem**. It is a simple classification technique, but has high functionality. Complex classification problems can also be implemented by using Naive Bayes Classifier.

**Gaussian Naïve Bayes Method**

Gaussian naïve Bayes formula is taken from Bayes theorem and given as

$$P(Y|X) = \frac{P(Y) \prod_{i=1}^n P(X_i|Y)}$$

P(X)

To calculate the numerator value using Gaussian distribution find mean and variance and substitute in the formula

$$P(X_i | Y) = \frac{1}{\sqrt{2\pi}\sigma_y} \exp\left(-\frac{(X_i - \mu_y)^2}{2\sigma_y^2}\right)$$

Where  $\mu$  is the mean and  $\sigma^2$  is the variance

Substituting the equation for the data set proposed classification is done for 3 classes namely advanced, middle, starting stage. With the 3 features extracted namely structural features, intensity of pixels, statistical features, data is tested and classification is done. Using the above formula the probability of three classes are calculated. Features are denoted by X and classes are represented by Y. The subset of X namely X1 stands for structure, X2 for intensity and X3 for statistics. The subset of Y namely Y1 stands for advanced class, Y2 for middle and Y3 for starting class.

Probability of advanced stage is

$$P(Y1|X) = \frac{(P(Y1)[P(X1|Y1)P(X2|Y1)P(X3|Y1)])}{P(Y1).P(X|Y1)+P(Y2).P(X|Y2)+P(Y3).P(X|Y3)}$$

Probability of middle stage is

$$P(Y2|X) = \frac{P(Y2)[P(X1|Y2)P(X2|Y2)P(X3|Y2)]}{P(Y1).P(X|Y1)+P(Y2).P(X|Y2)+P(Y3).P(X|Y3)}$$

$$P(Y1).P(X|Y1)+P(Y2).P(X|Y2)+P(Y3).P(X|Y3)$$

Probability of starting stage is

$$P(Y3|X) = \frac{P(Y3)[P(X1|Y3)P(X2|Y3)P(X3|Y3)]}{P(Y1).P(X|Y1)+P(Y2).P(X|Y2)+P(Y3).P(X|Y3)}$$

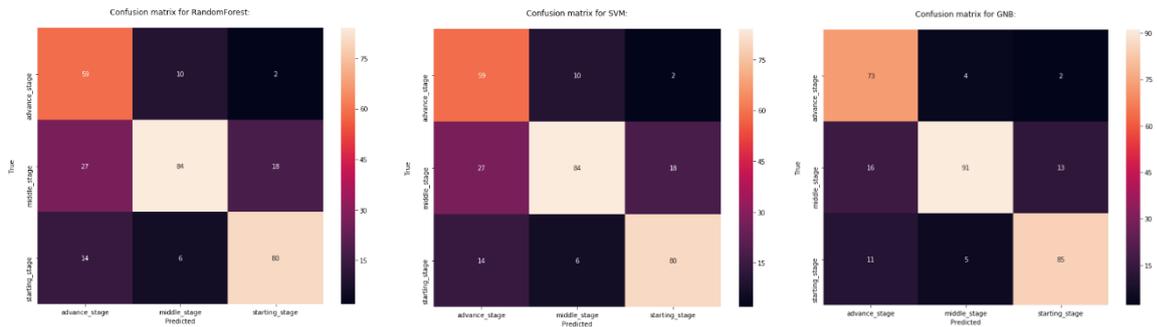
$$P(Y1).P(X|Y1)+P(Y2).P(X|Y2)+P(Y3).P(X|Y3)$$

Next mean and variance of X that is features values are calculated and substituted in the numerator value and result is derived and compared for highest value. For the sample image which class is giving the highest probability, the image belongs to the particular class.

**3.6. Experimental results for classification first method**

The result for classification methods discussed above are calculated using confusion matrix and the accuracy values are compared using all the three models for three stages of cancer. The confusion matrix for the

three models are given below in figure 3.6.1. From the confusion matrix sensitivity, specificity, accuracy, precision and F1 score are calculated for the three models.



Comparison of these values are given as tables 3.6.2.(a), (b), (c) below.

Figure 3.6.1. Confusion matrix for the three models

Table 3.6.2(a) Random forest result

STAGES	SENSITIVITY	SPECIFICITY	ACCURACY	PRECISION	RECALL	F1 SCORE
ADVANCED	82.09	83.09	84.33	61	84	71
MIDDLE	67.11	90.64	78.66	84	67	74
STARTING	80.01	90.02	88.66	81	80	81

Table 3.6.2(b) SVM result

STAGES	SENSITIVITY	SPECIFICITY	ACCURACY	PRECISION	RECALL	F1 SCORE
ADVANCED	83.09	82.09	82.33	59	83	69
MIDDLE	65.11	90.64	79.66	84	65	73
STARTING	80.01	90.02	86.66	80	80	80

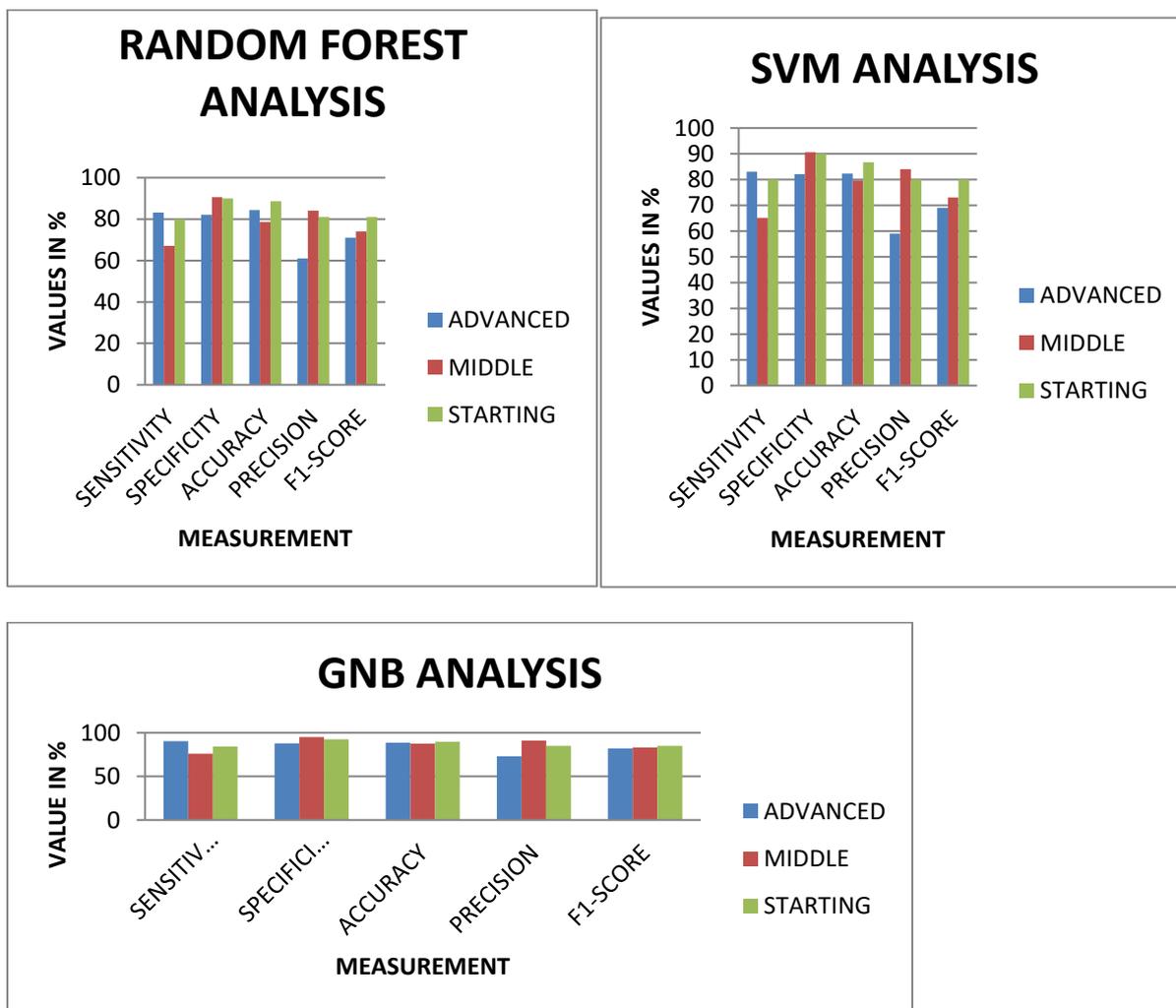
G						
---	--	--	--	--	--	--

Table 3.6.2(c) Gaussian Naïve Bayes result

STAGES	SENSITIVIT Y	SPECIFICIT Y	ACCURAC Y	PRECISIO N	RECALL	F1 SCORE
ADVANC E	90.12	87.78	88.41	73	92	82
MIDDLE	75.83	95	87.33	91	76	83
STARTIN G	84.15	92.46	89.66	85	84	85

The maximum accuracy from the three methods are given and GNB method shows higher accuracy of 85%. The bar chart for the given result analysis is given as figures 3.6.3.

Figure 3.6.3. Bar chart of three models

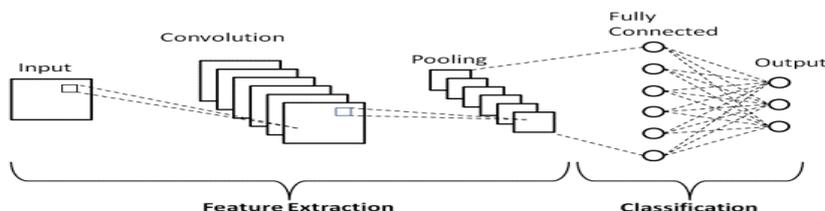


4. Classification using Deep Learning method

Using deep learning method thermal images are segmented using U-Net architecture and the results are already discussed in this paper. From the segmented images, features are extracted and classified using three deep learning models. They are CNN model, Inception V3 and Residual Network model.

Combination of neural network with a set of convolution operations is called Convolutional Neural Network model, a deep learning algorithm used for classifying the images especially in medical imaging. The general architecture of CNN is given below in figure 4.1.1

**4.1. CNN Model**



**Figure 4.1.1. General CNN architecture**

**BASIC STEPS FOR CLASSIFICATION**

1. Import packages and load the image
2. Split the data for training and testing purpose 80% for training and 20% for testing.
3. Construct a model using CNN architecture. The model contains 3 convolution layers with max Pooling layers with Global Max Pooling layer as the end with fully connected layer to classify the image whether they belong to class 0, 1, 2.
4. Class 0 stands for advanced stage, class 1 stands for middle stage and class 2 stands for starting stage.

**Figure 4.1.2 Architecture of CNN model**

NO. LAYER	OUTPUT SIZE	FILTER SIZE	STRIDE SIZE	DROPOUT
1. Input	200X200X3	----	----	----
2. ConvLayer	197X197X64	4X4	----	----
3. Relu	197X197X64	----	----	----
4. M.Pooling	98X98X64	----	2X2	98X98X64
5. ConvLayer	95X95X64	4X4	----	----
6. Relu	95X95X64	----	----	----
7. M.Pooling	47X47X64	----	2X2	47X47X64
8. ConvLayer	44X44X64	4X4	----	----
9. Relu	44X44 X64	----	----	----
10. M.Pooling	22X22X64	----	2X2	22X22X64
11. Global Max. Pooling	64			
12. Fully connected layer1	32			
13. Dropout				
14. Fully connected layer2	3			

The architecture contains 3 convolutional layer for extracting features and 3 max pooling layers to reduce the size of the input image. Dropout is used to remove unwanted parameters and global max pooling function is used to get flattened values which is given as input to

the dense layer which acts as fully connected layer. Two fully connected layers are used where the features extracted are combined to predict the classes. For three classes to predict three neurons are selected and given as output of the model using Artificial Neural Network

finally attached to the CNN architecture. Thus the model is created and trained using the data set. This model can be tested with any new image to predict which class it belongs that is advanced, middle and starting class.

#### 4.2. Inception V3 model

The next two models discussed in this paper are Inception V3 and Resnet 50 model based on transfer learning method. To create CNN model using transfer learning by these methods pre trained models are used which are created for similar problem and modification made to suit the current problem of identification. The same layers in the ImageNet architecture are used which is used to classify 1000 objects and finally the output layer is replaced to solve the current problem and the entire network is used as fixed feature extracted for the new data set. Architecture

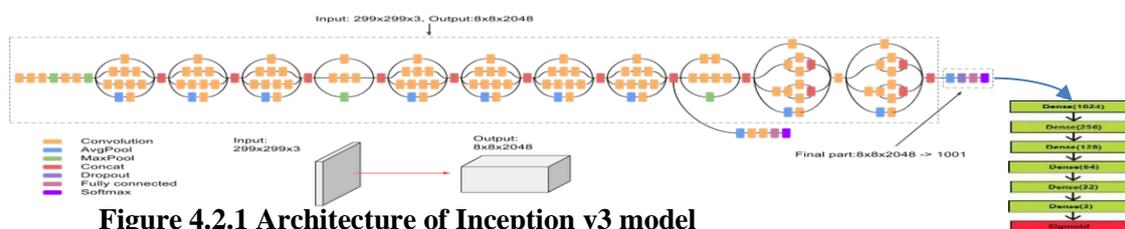


Figure 4.2.1 Architecture of Inception v3 model

The proposed architecture steps are as follows

1. The last layer which classifies 1000 classes is removed and proposed layer which classifies only 3 classes needed is inserted.
2. Next the inception v3 model is initialized.
3. The image size in the model is changed to the required size
4. Global max pooling function is used to get the number of features extracted to classify
5. Flattening and appending folder as dense layers in the last 6 layers produces the correct classification result as the image belongs to advanced, middle or starting stage.
6. Finally 1024 parameters are received which is reduced to 256 and 128 and 32 and finally only 3 parameters are selected for 3 classes
7. Sigmoid classifier is used to classify the images.

The model is created using training data set and used to classify any image based on this model to find whether the image belongs to advanced, middle or starting stage.

#### 4.3. Resnet50 Model

Another CNN model using transfer learning for classification of images with the Residual Network

The components used in this architecture are

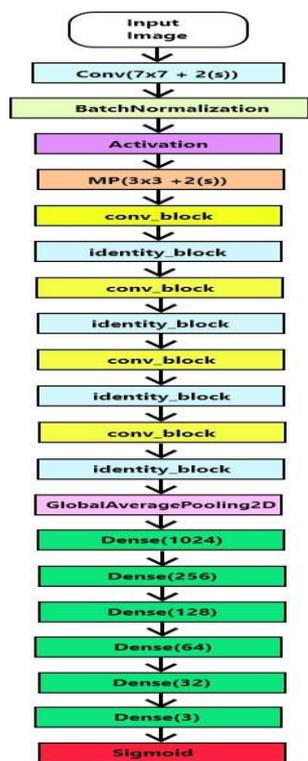
- Different filter size are used parallel to capture all the information
- Factorized the filter values to reduce the number of parameters to reduce the computational resources.
- Auxillary classifiers are used to prevent vanishing gradient problem.

This is called 48 layer network model used for image recognition and classification

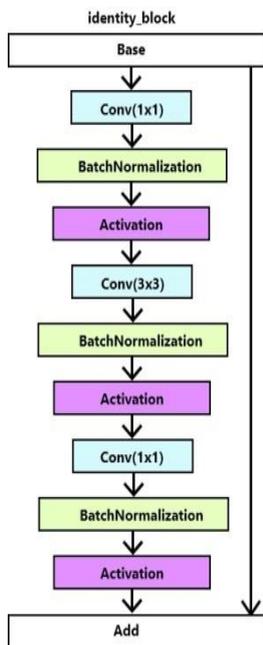
model is created using two functions called identity block and the convolution block. This model is created as the solution for the vanishing gradient problem which is defined as values in the layers when multiplied again and again with the derivative at the earlier layers becomes very small. The solution is given by Resnet model which uses skip connections to increase the weight by adding the actual input value say  $x$  along with the functions of  $x$  from convolution layer. The main logic of ResNet model is making both the input and output value equal. This is possible when  $F(X) = 0$  then  $Y = X$

In all algorithms data are trained on basis of  $Y$  that is output value but in Resnet data are trained on basis of  $F(X)$  value.

The proposed architecture is described with the basic building block of rows, columns, channels and classes. In this model skip connections allow the gradient to be directly back propagated to earlier layers. There are basic layers with  $7 \times 7$  convolution layer with stride 2 and the input image is given to this layer. First batch normalization and activation functions are performed. Next  $3 \times 3$  max pooling layer is performed with stride 2

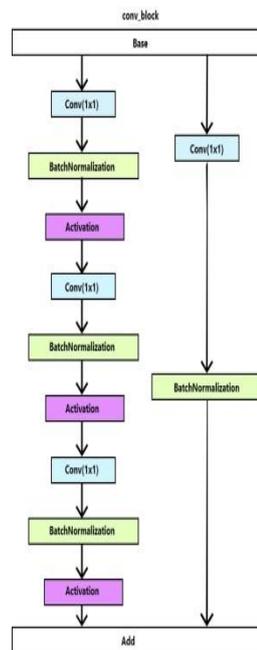


and then comes convolution block and identity block. In the proposed model there are 4 such operations with convolution block and identity block. In identity block the following steps take place. They are



In convolution block the steps are

1. on
- 2.
- 3.
4. The
- 5.



When input size is not equal to output Perform 1x1 convolution the base that is value from previous layer. Batch normalization is performed with activation. Additional 1x1 convolution layer is added in this block. The final steps are Finally global average pooling function is performed to get the feature values Then dense layers are added and the final value

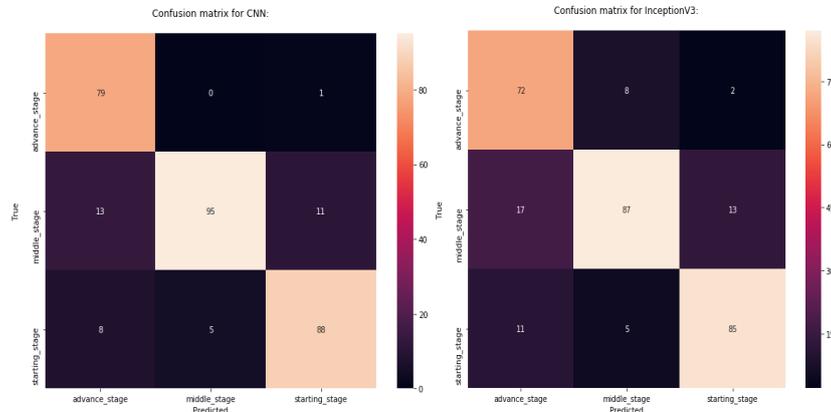
is given to sigmoid classifier for classification of three classes  
 6. The last layer functions are similar to Inception v3 model  
 Thus the model is created by changing the last layer and used for classification of thermal images.  
 The Resnet model is given in figure 4.3.1

1. The value from previous layer is taken as base
2. Convolution layer with 1x1 filter is performed followed by batch normalization and activation. Next 3x3 convolution is performed with batch normalization and activation and finally 1x1 convolution is performed with batch normalization and activation and add the final values to the next layer.

**Figure 4.3.1 Architecture of Resnet model**

**5 Experimental results**

The data set is trained with the three deep learning models and the comparison of the results are made with accuracy. The result obtained is shown in the form of confusion matrix to calculate the sensitivity, specificity, accuracy, precision and F1 score shown in figure 5.1.

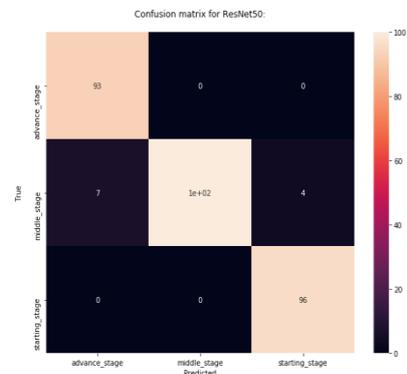


**Figure 5.1 Confusion matrix for 3 models**

All the three models are trained and Resnet50 model shows highest accuracy of 96% when training the thermal images. Performance analysis are done which are shown as table 5.2.(a),(b),(c)

**Table 5.2.(a) Result of CNN model**

STAGES	SENSITIVIT Y	SPECIFICIT Y	ACCURAC Y	PRECISIO N	F1 SCORE
--------	-----------------	-----------------	--------------	---------------	-------------



ADVANCED	<b>98.75</b>	<b>90.45</b>	<b>92.66</b>	<b>79</b>	<b>88</b>
MIDDLE	<b>79.83</b>	<b>97.23</b>	<b>90.33</b>	<b>95</b>	<b>87</b>
STARTING	<b>87.12</b>	<b>93.96</b>	<b>91.66</b>	<b>88</b>	<b>88</b>

**Table 5.2.(b) Result of Inception v3 model**

STAGES	SENSITIVIT Y	SPECIFICIT Y	ACCURAC Y	PRECISIO N	F1 SCORE
ADVANCED	<b>87.80</b>	<b>87.15</b>	<b>87.33</b>	<b>72</b>	<b>79</b>

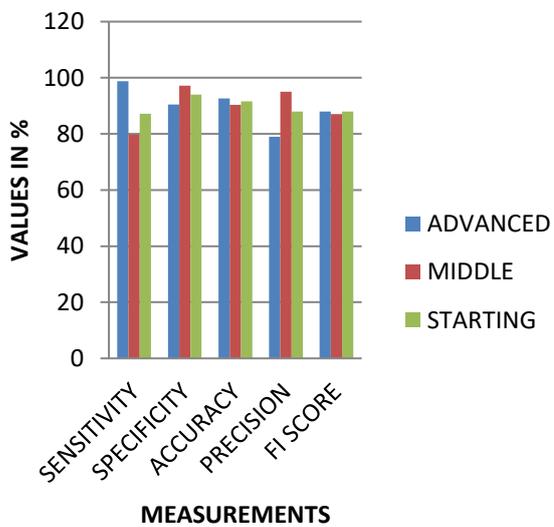
MIDDLE	74.35	92.89	85.66	87	80
STARTING	84.15	92.46	89.66	85	85

Table 5.2.(c) Result of Resnet50 model

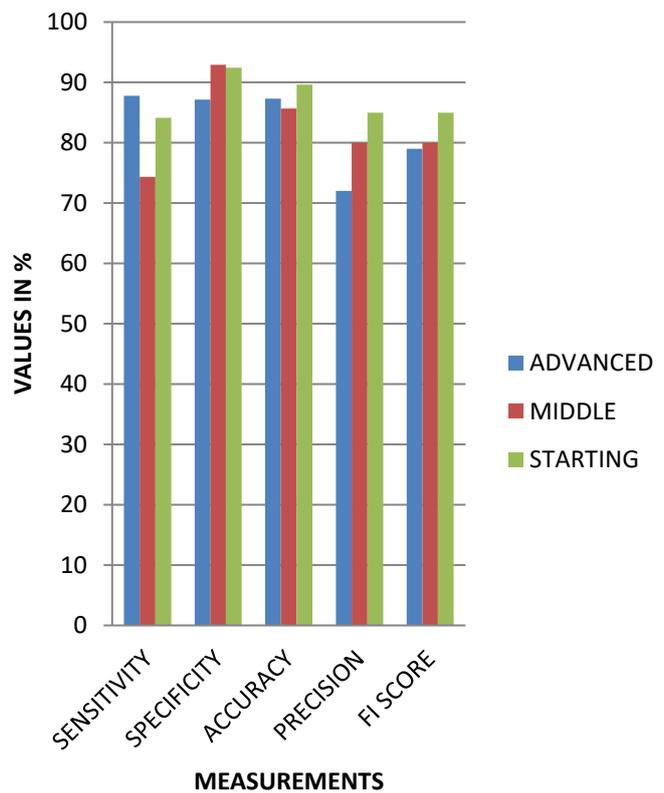
STAGES	SENSITIVIT Y	SPECIFICIT Y	ACCURAC Y	PRECISION	F1 SCORE
ADVANCED	99.99	96.01	97.66	93	96
MIDDLE	90.09	99.98	96.33	100	95
STARTING	99.00	98.03	98.66	96	98

For the above result bargraphs are constructed shown in figure 5.3

### CNN METHOD



### INCEPTION V3



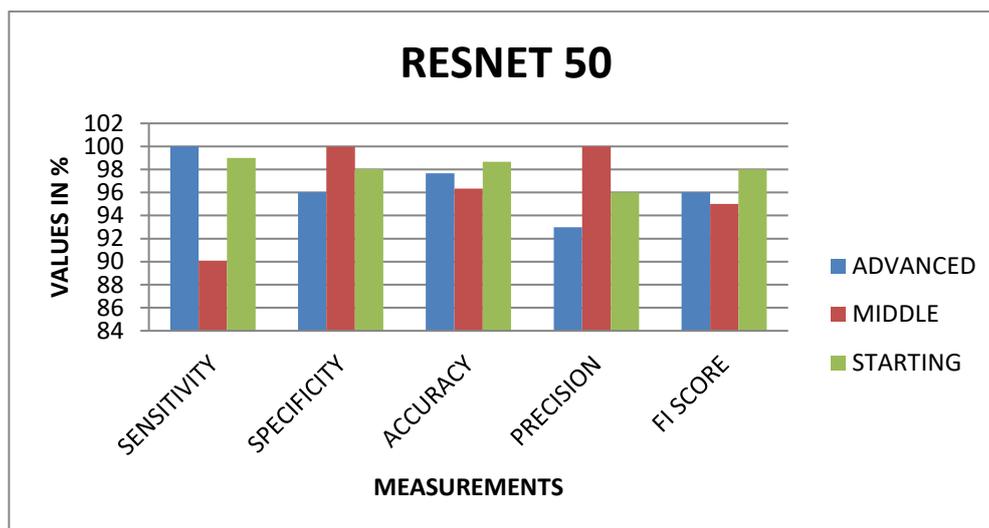


Figure 5.3 Barcharts of three models

## 6 Conclusion

Finally the thermal images are collected and segmented using two novel approaches to divide the image into normal and abnormal. Features are extracted from the abnormal image using three methods namely structural, histogram and statistical method. The features are used to classify the segmented image into three classes called advanced, middle and starting stages. The classification methods used are Random forest, Support Vector Machine, Gaussian Naïve Bayes model along with deep learning models like CNN mode, Inception V3 and resnet50. Among these models resnet50 model shows accuracy upto 96% whereas Gaussian naïve model showed 83% accuracy comparatively.

## References

- Halperin E, Perez C, Brady L, Ed. Perez and Brady's Principles and Practice of Radiation Oncology: Fifth edition. Lippincott Williams & Wilkins, Philadelphia: 2008. 1162-1317. [http://www.cancervic.org.au/images/content/breast\\_health/lcis.gif](http://www.cancervic.org.au/images/content/breast_health/lcis.gif)
- Breast Cancer Detection in Mammograms using Convolutional Neural Network Saira Charan, Muhammad Jaleed Khan, Khurram Khurshid 2018 International Conference on Computing, Mathematics and Engineering Technologies-iCoMET 2018 978-1- 5386-1370-2/18/\$31.00 ©2018 IEEE
- Development of a Deep-Learning-Based Method for Breast Ultrasound Image Segmentation Rania Almajalid<sup>1,2</sup>, Juan Shan<sup>1,\*</sup>, Yaodong Du<sup>1</sup>, Ming Zhang<sup>3,4,\*</sup> 2018 17th IEEE International Conference on Machine Learning and Applications
- Breast cancer detection using deep convolutional neural networks and support vector machines Dina A. Ragab<sup>1,2</sup>, Maha Sharkas<sup>1</sup>, Stephen Marshall<sup>2</sup> and Jinchang Ren<sup>2</sup>
- Region based segmentation by the generation of hierarchical code book using thermal images P.Kanimozhi, S.Sathiya, Solidstate technology .us
- Face detection based on invariant moments classified by neural network Milan Tuba, Romana Capor-Hrosik, and Mirjana Vukovic
- An Efficient Algorithm for Recognition of Human Actions Yaser Daanial Khan,<sup>1</sup> Nabeel Sabir Khan,<sup>1</sup> Shoaib Farooq,<sup>1</sup> Adnan Abid,<sup>1</sup> Sher Afzal Khan,<sup>2</sup> Farooq Ahmad,<sup>3</sup> and M. Khalid Mahmood<sup>4</sup>
- FEATURE EXTRACTION OF THERMAL IMAGE FOR BREAST CANCER ANALYSIS 1P.Gomathi, 2V.Jamuna 1,2Research Scholar M.Kumarasamy College of Engineering, Tamil Nadu, India 1 gomathipr.ece@mkce.ac.in, 2 jamunav.ece@mkce.ac.in
- Breast Cancer Detection, Segmentation and Classification on Histopathology Images Analysis: A Systematic Review R. Krithiga<sup>1</sup> · P. Geetha<sup>1</sup>
- Deep Convolutional Neural Networks-Based Automatic Breast Segmentation and Mass

Detection in DCE-MRI Han Jiao , 1Xinhua Jiang,2 Zhiyong Pang , 1Xiaofeng Lin,2 Yihua Huang , 1 and Li Li

11. Involvement of Machine Learning for Breast Cancer Image Classification: A Survey Abdullah-Al Nahid and Yinan Kong
12. Classification of breast cancer histology images using Convolutional Neural Networks Teresa Araujo<sup>1,2</sup>\*, Guilherme Aresta<sup>1,2</sup>, Eduardo Castro<sup>1</sup>, Jose´ Rouco<sup>2</sup> ,
13. A Study of Fine-Tuning CNN Models Based on Thermal Imaging for Breast Cancer Classification Roslidar Roslidar<sup>1,2</sup>, Khairun Saddami<sup>1</sup> , Fitri Arnia<sup>2</sup> , Maimun Syukri<sup>3</sup> , Khairul Munadi<sup>2</sup>\* 2019 IEEE International Conference on Cybernetics and Computational Intelligence (IEEE CYBERNETICSCOM),