iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** V **Month of publication:** May 2018

DOI: http://doi.org/10.22214/ijraset.2018.5002

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue V, May 2018- Available at www.ijraset.com

Detecting Melanoma Cancer using Image Processing

Vaishnavi. K.R[1], Sheethal. K[2], Manjunath C.R[3]
[1, 2, 3]*Department of Computer Science and Engineering, Jain University, Bangalore, India*

Abstract: ***Skin cancer is the most dangerous form of cancers. The most common form of skin cancer is Malignant melanoma. If this cancer is detected in the early stage, then it can be cured easily. In this paper we present an overview of all the classification methods that are used to detect the skin cancer and a comparative study of all the discussed classification methods. Image classification depends on various factors and it is a complex and vital process. Image classification uses both supervised and unsupervised classifications and it is considered as one of the best method to predict the skin cancer results.***
Keywords: Skin cancer, classification methods, Support Vector System, Decision Trees, Artificial Neural Network, K-nearest neighbor

I. INTRODUCTION

Skin is an important part of the human body. Skin protects us harmful elements such as ultra violet rays and help to maintain the body temperature. Malignant melanoma is the most often type of skin cancer and it begins in skin cells called melanocytes and the disease can start growing from anywhere on the body. They appear in different colors such as brown, black, red, blue and combination of blue and red. Melanoma skin cancer can be caused only on the top layer of the skin. The tumor is much less than 1mm-2mm thick and the surface may or may not be broken. The cancer cells can spread to other tissues, liver, brain, bones and other organs. Classification difficult and complex task. In classification –image classification is the best suited for detecting the skin cancer. Image classification is a method which contains a database consisting of predefined patterns which will help us to compare with different objects and then group according to a specific category. Classification technique can be used in various applications such as remote sensing, robot navigation, biomedical images.

Advantages of classification are:

1) It can be used for wider applications.
2) The results are more consistent than any other methods.
3) Has the necessary capacity to compare as many objects as possible.

Different stages of classification image sensing, image pre-processing, segmentation, feature extraction, classification. After classification methods are applied on the skin, we will be able to determine whether the skin sample is cancer-type or non-cancer-type.

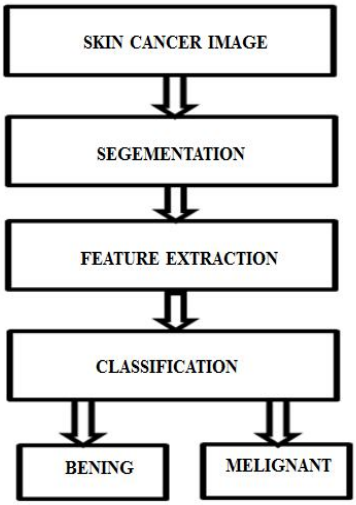


Figure.1.1

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue V, May 2018- Available at www.ijraset.com

II. BACKGROUND WORK

Classification of images into cancer-type and non-cancer-type can be done with the help of image classification. Image classification involves two rule-based classification algorithms: - supervised learning and unsupervised learning. We need to have complete knowledge and then proceed for testing and this information is gathered by the analyst. In supervised learning, we will be able to predict the result or class of data and a trained data set will be provided. When we are implementing supervised learning method, prior information is essential before testing and this data must be gathered by the analyst.
Supervised classification involves the following steps:

1) Identify the training areas for each class.
2) Identifying the signatures such as mean, variance, co- variance.
3) Classifying the entire pixels.
4) Mapping of information to a particular class.

In Supervised classification the user can identify the errors and will be able to rectify them. But Supervised classification is time consuming and costlier and the training data is selected by the analyst may not involve all the conditions to detect the skin cancer. Supervised classification also involves the human intervention.
In unsupervised learning, no trained data will be provided but the classifiers itself have to find to which category or class it belongs to. The user need not have any prior information that is no human intervention is required. It involves the following steps:

1) The collected data must be clustered
2) Classifying the pixels based on the clusters.
3) Spectral class map.
4) The analyst must label the cluster.
5) Mapping of information to a particular class.

Unsupervised classification is much faster when compared to supervised classification. This method is free from human errors and there is no requirement of prior knowledge. In supervised classification parametric and non-parametric classification plays a major role in detecting melanoma skin cancer.

A. Parametric classifier

In parametric classifier we mainly use mean vector and covariance matrix. The major disadvantage of parametric classifier is that it generates the noisy results and it is difficult to integrate spatial. Few examples of parametric classifier are maximum likelihood method and linear discriminate analysis.

B. Non-Parametric classifier

This classifier gives better results when compared to parametric classifier. Here, for performing the class separation we don't use statistical parameters.
Few examples of non-parametric classifier are: Expert system, support vector machine artificial neural network, decision tree classifier. In this paper we are mainly discussing about parametric classifiers and non- parametric classifiers. Parametric classifiers involve decision tree, support vector system. Non- Parametric classifiers involve artificial neural network and k-nearest neighbor.

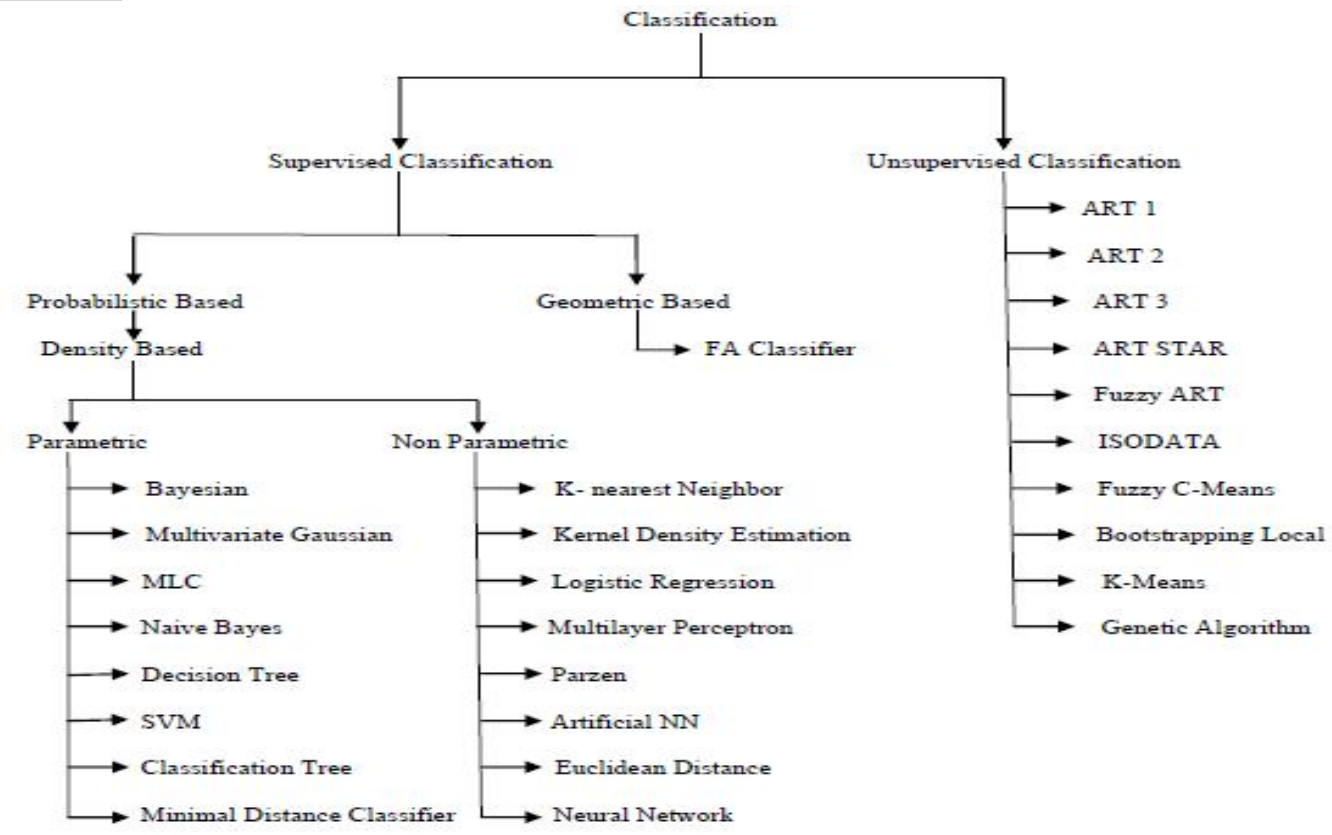


Figure1.2

III. PARAMETRIC CLASSIFIER

A. Support Vector Machine:

Support vector machine is one of the methods of supervised learning, which will help to analyze the data that is used for the classification.

1) Advantages of Support vector machine:
a) It has the ability to give a unique solution.
b) More efficient than other methods.
c) Avoid over-fitting

2) Disadvantages of SVM are:
a) High algorithm complexity
b) Run slowly

SVM is again classified as linear SVM and non- linear SVM.
Linear SVM- In this method, for a given dataset of n points find the maximum-margin hyperplane by dividing the points into groups. It is again divided into hard- margin and soft- margin. Non-linear SVM- create nonlinear classifiers by applying the kernel trick to maximum-margin hyperplanes.

B. Decision Tree

The decision tree approach belongs to the supervised classification. Decision trees are like a tree like structure. Each branch represents the decisions to be made graphically. Decision trees algorithm constantly divide the given data set according to a particular criterion, resulting in a tree-like structure. Here each branch represents the decision to be made graphically. The decision tree

algorithm follows rule-based classification. This method allows the acceptation and rejection of class label at every intermediate stage. Decision trees gives the set of rules after classification that should be understood.

1) Advantages Of Decision Tree

1) Does not require much effort from the users.
2) It has easy interpretation.
3) Simple representation.
4) Efficient use of decision making.
5) They are not black-box models but can easily be expressed as rules.

2) Disadvantages Of Decision Tree

a) Splits are very sensitive to training data set
b) High classification error rate.
c) Greedy construction process that is, in each step the best split has to be selected to obtain better results.

Decision tree algorithm is widely on various kinds of domain, such as machine learning, artificial intelligence (AI), pattern recognition and data mining. A decision tree algorithm is a flow-chart-like tree structure. The internal nodes represent the attribute and the branch node represents the test outcome. The main advantage of this method is that, one can easily understand the logic involved in it and its representation is very simple. The decision tree algorithm is not well suited for this domain. Using Naïve Bayesian Classification is effective model for diagnosis Malignant Melanoma, but the decision tree algorithm is not well suited for this domain.

IV. NON-PARAMETRIC CLASSIFIER

A. Artificial Neural Network

Artificial neural network is an efficient computational model based on its structure and functions of biological neural networks. In artificial neural network the structure is affected when Information flows through the network as because a neural network changes based on the input and output. Complex relationships between inputs and outputs are modelled or patterns are found where artificial neural network is considered nonlinear statistical data modelling tool.

Artificial neural network contains numerous modifications. The simplest of those is in consideration with static types have one or more static components, including the number of units, number of layers and unit of weights and topology. Some types operate independently, while others allow learning to be supervised by the operator

An artificial neural network visually or textually goes through a training various phase where it learns to recognize patterns in data. The network compares its actual output produced with what actually meant to produce, that is the desired output during the supervised phase. Using the backpropagation difference between both outcomes is adjusted. In order to adjust the weight of its connections the network works backwards going from the output unit to the input unit that is between the units until the difference between the actual and desired outcome produces the lowest error as much as possible.

To automate the design of Artificial Neural Network (ANNs), Neural architect search (NAS) uses machine learning. Various approaches to neural architect as they have designed networks that compare well with hand -designed systems. The aim of a basic search algorithm is to propose a candidate model to evaluate it against a dataset and use the results as feedback to teach the Neural Architect Search(NAS).

1) Advantages of Artificial Neural Network: -

a) This network is easy to use
b) Efficient for complex problem like image recognition.
c) Regardless of its linearity can approximate any function.

2) Disadvantages of artificial neural network: -

a) Linear regression would be best solution for often cases.
b) Enhancement of accuracy by few percent can increase the scale by several magnitudes.
c) Requires increased training and test cases.

3) Characteristics of Artificial Neural Networks

a) Choice of model: This model is very complex with slow learning process. As the focus is mainly on the data representation and the application.

b) Learning algorithm: Between the learning algorithms there are various trade-offs. As most of the algorithm will function well with the actual perfect hyperparameters of training on a particular data set aspect. Therefore, tuning and selecting an algorithm for training on undiscovered data desires symbolic experimentation.

c) Robustness: If the preferred model's cost function and learning algorithm are named appropriately, the resulting Artificial neural network can become potent.

In practical applications artificial neural network plays role in email service, as to find and eliminate spam from end user's inbox-commerce rostrum that are being used to illustrate reconditions to their audience, chatbots are being developed with artificial neural network as for numerous universal language processing and mainly in detection of skin cancer diseases.

B. K- Nearest Neighbor

K-nearest neighbor is considered a non-parametric classifier, maned as lazy learning algorithm. To predict classification of a new sample point where the purpose is to use a database in which the data points are separated into several classes. As K-Nearest Neighbor doesn't learn a descriptive function from the training data but "memorizes" the training dataset instead hence it's a lazy learner. One of the simplest of all machine learning algorithms is k-nearest neighbor. The K-Nearest Neighbors (K-NN) is a non-parametric method being used for regression in pattern recognition and as well as classification aspect. In both of the aspects, the input consists of the k-closest training examples in the feature space. The output being mainly focused on whether K-Nearest Neighbor is pre-owned for classification or regression.

1) In K-NN classification, the outputs are considered as a class association. Classification is being done on object by a majority vote of its neighbors, as objects being assigned to the class and being most common among its k-nearest neighbor. The object is assigned to the class of that single nearest neighbor of the value of k=1.

2) The output is considered as the property value for the object in k-nearest neighbor regression. The value is the average of the values of its k nearest neighbors.

The neighbors are taken from asset of objects for which the class or the object property value is known. This can be thought of as the training set for the algorithm, though no explicit step is required. K-Nearest Neighbor algorithm has a peculiarity feature is that it is sensitive to data. One more algorithm named k-means not being confused, which is another popular machine learning technique.

1) Few other features of KNN

a) K-Nearest Neighbors stores the complete training dataset and it uses for its representation.

b) No learning model is needed for K-Nearest Neighbors.

c) It has the ability to calculate the time between the input sample and training set.

2) Advantages of K-Nearest neighbors

a) Insensitive to outliers -accuracy can be affected from noise or irrelevant features

b) No assumptions about data - useful, for example, for nonlinear data

c) Simple algorithm – to explain and understand.

d) High accuracy (relatively) - it is pretty high but not competitive in comparison to better supervised learning models

e) Versatile - useful for classification or regression

3) Disadvantages of K-Nearest Neighbors

a) High memory required is high.

b) As the algorithm stores all of the training data hence it is computationally expensive.

c) State of Prediction might be slow (with big N).

d) Its Sensitive to irrelevant features and the also scale of the data.

4) Summary of the K-neighbor algorithm

a) Specifying the k integer, with a new sample.

b) Selecting the k entries which are closest to the new sample.

c) Finding the common entries.
d) Resultant is the new classification.

5) Applications of K-Nearest Neighbors
a) Based on clinical and demographic variables Identification of the risk factors for prostate cancer was being done.
b) For text categorization or text mining K-Nearest Neighbor algorithm is one of the most popular algorithms
c) One of the most core financial tasks of K-Nearest Neighbor is Stock market forecasting.

6) Applications of K-Nearest Neighbor in finance are includes
a) Trading futures
b) Credit rating
c) Loan management
d) Bank customer profiling
e) Money laundering analyses

V. LITERATURE SURVEY

[4] Have proposed an automated system to detect the skin cancer using multi class Support vector system. The accuracy of this proposed system is 83.43%. By using Gray Level Co-occurrence matrix (GLCM) different features such as energy, entropy, etc. are extracted. Multi class support vector machine with the combination of multiple binary classifiers are used to train and classify the random input images. By using multiclass support vector machine, one can detect multiple cancers with better accuracy and performance.

[5] Proposed a system that uses Support Vector System (SVM) using Principle Component Analysis (PCA) and Grey Level Co-occurrence Matrix (GLCM) is used for feature extraction. By extracting features from PCA, they were very much successful in classifying the extracted lesion region of interest (ROI). By following this method, they were able to get 92.1% of accuracy. This accuracy was achieved by selecting 11 features of confusion matrix and the same accuracy was observed by selecting 5 features using PCA. From these results we can easily determine whether the collected skin sample is cancerous or non-cancerous. By using SVM and PCA we will be able to detect the skin cancer not only for 11 or 5 features but even for the complete set of features (with lower computational complexity). This author has proposed a very good method which gives accurate results when compared to other methods. By using this idea, we can propose a system that uses a mobile application to predict the results.

[6] This author proposes a system a novel hierarchical classification system. The accuracy of the proposed system to determine whether the skin cancer is cancerous or non–cancerous is 94% but in detecting the melanoma skin cancer the accuracy is 74%. Hierarchical classification has more flexibility and efficiency. In this proposed system the main classification is divided into smaller classification. From these smaller nodes, one can be able to determine whether the skin sample is cancerous or non- cancerous.

[8] Uzmabano Ansari, Tanujasarode proposes an easy method rather than time consuming laboratory method. This method is painless and not time consuming. In image preprocessing noise removal and image enhancement, in image segmentation thresholding method is used and for classification SVM is used. In this method 95% of accuracy is achieved.

In this proposed system, first the impost image is collected then the image is pre-processed by three main things that is noise removal, gray scale conversion and image enhancement followed by segmentation, feature extraction and classification.

Skin cancer detection is done using Support Vector Machine (SVM). Gray Level Co-occurrence Matrix (GLCM) is used to extract the features from the image, and then to determine whether the image is cancerous or non- cancerous. The purpose of feature extraction (glum) is to suppressed the original image data set by measuring certain values or features that helps to classify different images from one another.

PARAMETERS	SVM CLASSIFICATION
TRUE POSITIVE	16
TRUE NEGATIVE	3
FLASE POSITIVE	0
FLASE NEGATIVE	1
ACCURACY	95%

[11] Samy Bakheet, proposed a system by using CAD (Computer-Aided Diagnosis), it is an effective framework developed on the basis of Support vector system model on an optimized set of HOG (Histogram of Oriented Gradient). By using this method, they were able to get 97. 32% accuracy.

In this paper, the classification model used is supervised learning (which includes some). The collected skin lesion is mainly divided into two classes and then determining whether the skin sample is cancerous or non-cancerous. Pattern recognition, with SVM is used to characterize substantial resistance to overfitting, inherent problem for several supervised learning algorithms and a long-standing.

The main idea is to build two parallel hyperplanes. Then, SVM tries to find the separating hyperplane that maximizes the distance between the two parallel hyperplanes.

In the presented approach, two classes of skin lesions are created i.e. an SVM classifier with the radial basis function (RBF) kernel and default parameters the major benefit of using an RBF kernel over the other kernels is that linear kernel has limits data set and specified boundaries. It also has fewer numerical difficulties. During the test phase, a test unseen skin lesion sample is given to the trained SVM model for classification, based on the extracted features.

The proposed system is verified in terms of k-fold cross validation. In this validation process two independent subset are chosen, that is, the training set and the test set. For four-cross validation ¾ of all images are used for training set and ¼ of images are used for testing set.

HOG features computation, which is considered as the best methods, will extract the skin lesion and is converted into gray scale image level and it is resized into 64X64 pix. To evaluate the performance the terms used here are sensitivity (SN), specificity (SP) and accuracy (AC).

$$SP = \frac{TN}{TN + FP} \times 100\,(\%).$$

$$SN = \frac{TP}{TP + FN} \times 100\,(\%).$$

$$AC = \frac{TP + TN}{TP + TN + FP + FN} \times 100\,(\%)$$

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

PARAMETERS	SVM CLASSIFICATION
TRUE POSITIVE	55
TRUE NEGATIVE	54
FALSE POSITIVE	2
FALSE NEGATIVE	1
SENSITIVITY	98.21%
SPECIFICITY	96.43%
ACCURACY	97.32%

[12] A Computer aided skin cancer detection system can achieve a new discovery of detecting benign or malignant skin lesions and separating them from healthy skins. The Artificial Neural Networks for the classification of Malignant Melanoma from benign melanoma. Dermoscopic images were collected and they are processed using median filter are used to remove salt and pepper noise. After preprocessing images is segmented using maximum entropy method. This Methodology has got 86.66% accuracy. The proposed method trained with 75% and tested with 25% of the total number of images. At the end of the training process updated weight values are stored. Then the performance value is measured. By varying the Image processing techniques and training

algorithms of Artificial Neural Network, the accuracy is improved for this system and the images are classified as cancerous or non-cancerous.

VI. METHODOLOGY

A. *Existing system.*

In this present era, there are many different methods to determine the skin cancer. Few methods are by using artificial neural networks, support vector machine as one of those in image processing. The above methods are not efficient in term of detecting unpredictable skin cancer.

B. *Proposed System*

In this paper we compare all the all the four classification methods with its advantages and disadvantages.

CLASSIFICATION METHODS	ADVANTAGES	DISADVANTAGES
Support Vector system	1. Deliver unique solution. 2. Very efficient than other methods. 3. Avoid over-fitting	1. High algorithm complexity 2. Run slowly
Decision Trees	1. Require little efforts from users 2. Easy to interpret and explain	1. High algorithm complexity 2. Run slowly 1. High algorithm complexity 2. Run slowly
Artificial neural network	1. Robust to noisy training dataset 2. Very efficient for large dataset	1. High computational cost 2. Lazy learner
K-Nearest neighbors	1.Insensitive to outliers -accuracy can be affected from noise 2. No assumptions about data - useful. 3. Simple algorithm - to explain and understand. 4. High accuracy 5. Versatile - useful for classification or regression	1.Computationally expensive - because the algorithm stores all of the training data 2. High memory requirement. 3. Stores all of the training data. 4. Prediction stage might be slow (with big N). 5. Sensitive to irrelevant features.

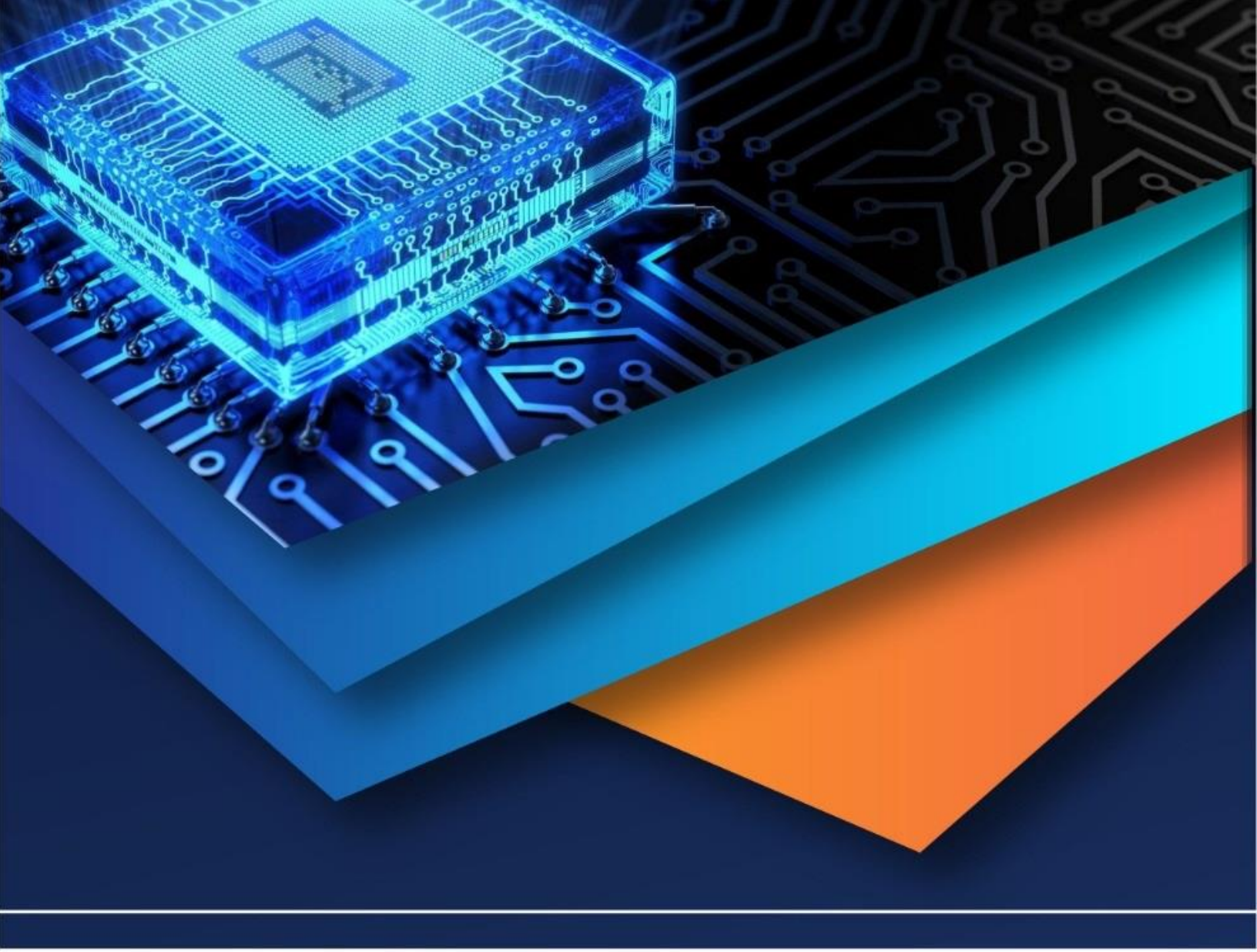
Table 4.1

VII. CONCLUSION

In image classification we can use both supervised and unsupervised methods to detect the skin cancer. In parametric, support vector system is best suited to detect the skin cancer and in non-parametric method, artificial neutral network is best with respect to accuracy aspects.

REFERENCES

[1] Chaitali Dhaware, Mrs. K. H. Wanjale , Survey on Image Classification Methods in Image Processing, (2016). International Journal of Computer Science Trends and Technology (I Jcs T) – Volume 4 Issue 3, May - Jun 2016
[2] Poornima M , Dr. Shailaja K, Detection of Skin Cancer Using Svm, (2017), International Research Journal of Engineering and Technology (Irjet) Volume: 04 Issue: 07 | July -2017.
[3] Sujaya Saha, Dr. Rajat Gupta, An Automated Skin Lesion Diagnosis by Using Image Processing Techniques, (2015), International Journal on Recent and Innovation Trends in Computing and Communication Issn: 2321-8169 Volume: 2
[4] Ms. H. R. Mhaske, Mrs. D. A. Phalke, Melanoma Skin Cancer Detection and Classification Based on Supervised and Unsupervised Learning,
[5] Hiam Alquran, Isam Abu Qasmieh, Ali Mohammad Alqudah, Sajidah Alhammouri, Esraa Alawneh, Ammar Abughazaleh, Firas Hasayen, The Melanoma Skin Cancer Detection And Classification Using Support Vector Machine, (2017), Ieee Jordan Conference on Applied Electrical Engineering and Computing Technologies (Aeect),
[6] Prashant Bhati, Manish Singhal, Early Stage Detection and Classification of Melanoma, 2015 International Conference on Communication, Control and Intelligent Systems (Ccis).
[7] Shivangi Jain, Vandana Jagtap, Nitin Pise, Computer Aided Melanoma Skin Cancer Detection Using Image Processing, (2015), International Conference on Intelligent Computing, Communication & Convergence (Iccc-2015).
[8] Uzmabano Ansari, Tanujasarode, Skin Cancer Detection Using Svm, (2017), Proceedings of Wrfer International Conference, 16th April 2017, Pune, India.
[9] Diwakar Gautam, Mushtaq Ahmed, Melanoma Detection and Classification Using Svm, Based Decision Support System, (2015), Ieee Indicon 2015 1570185633
[10] Siddhartha Sankar Nath, Jajnyaseni Kar, Girish Mishra, Sayan Chakraborty, Nilanjan Dey, International Conference on Control, Instrumentation, Communication and Computational Technologies (Icccict).
[11] Samy Bakheet, An Svm Framework for Malignant Melanoma Detection Based on Optimized Hog Features, (2017)
[12] Lucia Ballerini, Robert B. Fisher, Ben Aldridge, Jonathan Rees, 2014,



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY