iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VI **Month of publication:** June 2021

DOI: https://doi.org/10.22214/ijraset.2021.35285

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: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

Skin Cancer Diagnosis using SP-SIFT and Machine Learning

Miss. Divya Rajendra Dhamdhere[1], Dr. S. K. korde[2]
[1, 2]*M.E Computer Eigg. Dept, P.R.E.C. Loni, Ahmednagar, Maharashtra-413736., Savitribai Phule Pune University.*

Abstract: ***Today information handling is become an integral part of our day-to-day life as it is available in abundance. The information from the internet can be used for various purposes from Security to Healthcare management. It can decrease the load on a human being to handle the data manually. So, to highlight the importance of information handling to the society we choose medical field where images are in abundance on the internet. The analysis of images for decision making in medical perspective can be a boon to the medical field and decrease the work of the doctor and increase his productivity to a new level. The cancer has become a menace these days and has become a major health problem. Computer-aided diagnosis (CAD) systems can be used to provide an insight to the cancer specialist and help him to diagnose the various stages of cancer using images and CAD. So, we thought of designing a CAD system which will help in cancer detection and handling various stages of cancers using cancer images as the backend of our application. We propose to implement it on skin cancer which is an ever-increasing cancer due to pollution and environment changes. To design a successful CAD system for skin cancer, it has to be a combination of image processing and machine learning technologies together. In our proposed system we are going to accumulate a lot of images from various cancer image databases which are available free on the internet and design a cancer database with images put under various stages of skin cancer. After database we will apply SP-SIFT algorithm on it and extract features. This database has to available remotely so we have to maintain this feature set on the cloud. The extracted features will be used to train and get results from SVM and Naïve Bayes. Thus, our results will be divided in three categories melanoma, non-melanoma and benign which define stages of skin cancer. Thus, our system will help and improve the decision making and productivity of a medical practitioner.***

Keywords: Skin Disease, Image Processing, Machine Learning, SP-SIFT, SVM, Naïve Bayes, Image Database.

I. INTRODUCTION

Cancer is a collective term for a large group of diseases that can spread to any body part. One important characteristic which makes cancer deadly is the rapid creation of abnormal cells that grow beyond their usual limitations and become lethal, and then attack adjoining cells and spread to other organs. The term used for this process is metastasizing. According to World Health Organization Website Cancer is the second leading cause of death globally, and is responsible for an estimated 9.6 million deaths every year. 1 in every 6 deaths in the world is caused due to cancer as per the WHO data. As with any type of cancer, the earlier it is detected, the greater is the chance for survival. The cancer has two stages benign and malignant. In benign stage the cancer is curable and in malignant stage there are high chances of patient succumbing to cancer. For cancer diagnosis the healthcare professional has to depend on his viewing and testing various symptoms which can be time consuming. If a cancer is detected in early stages, it can be treated and has a high chance of survival then the patients where the cancer is detected late. The process of existing form of cancer diagnosis and treatment is demonstrated in the Fig.1.

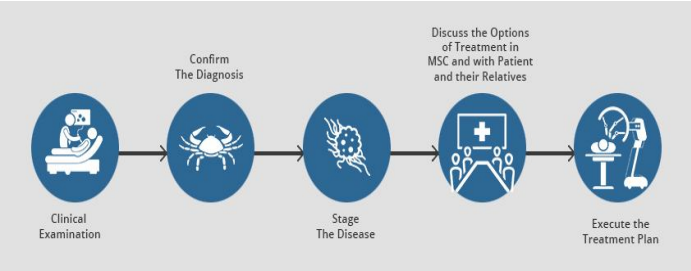


Fig. 1 Existing Cancer diagnosis and treatment process.

Thus, this kind of cancer diagnosis and treatment as shown in Fig.1 can be time consuming and costly as most of cancers are diagnosed at a late stage where a patient have a very low chances of survival. So there arises a need to develop a software which will help a medical professional and patient to analyze the symptoms using image processing and machine learning techniques which will help in finding the cancer stages easily. If cancer found at an early stage will be a boon for the patient to live a normal life again. So, in other words the main objective of this paper is to:

A. Focus on saving life of cancer patient by diagnosing the cancer at an initial stage.
B. To help medical practitioner diagnose a cancer without a time-consuming effort.
C. Study various image processing techniques that can be used in cancer detection.
D. To get a more accurate result, machine learning has to be used with image processing.
E. Propose a new fusion frame work with a combination of image processing, machine learning and cloud computing together.
F. Create and maintain a feature set database of various stages of cancer on cloud so that it can be accessed from anywhere using internet.
G. Evaluate and analyze the new skin cancer diagnosis framework and its strengths.

Thus, the rest of the paper is structured as follows:
1) Section II. explains literature survey which studies various techniques with their advantages and drawbacks.
2) Section III. explains the methodology i.e., mathematical model and algorithms to be used by the system.
3) Section IV. explains proposed system with block diagram or system architecture and working of the system.
4) Section V. shows the results and discussions and how it will be implemented.

II. LITERATURE SURVEY

This section describes the fundamentals of various techniques that can be used in designing a skin cancer diagnosis system. It helps in understanding various ideas put forward by various technical papers published by various authors and how they put forth a more accurate and secured techniques. Some of the ideas with technique and drawbacks are mentioned below:

In 2019 Navarro et al. [1] presented the paper focusses mainly on skin cancer diagnosis using SP-SIFT. This technique is quite good and covers all the things needed for a successful skin cancer detection system. But main drawback of this system is that it concentrates on feature extraction, Segmentation and not on prediction using machine learning where machine learning improves accuracy of the system.

In 2019 Guo et al. [2] presented the paper focusses mainly on mainly on image detection using SIFT algorithm. This technique is quite good and covers all the things needed for detecting and labeling images. But main drawback of this system is that it concentrates on feature extraction and labeling of images and not on prediction using machine learning where machine learning improves accuracy of the system.

In 2018 Yawen et al. [3] presented the paper focusses mainly on vehicle image detection using SIFT algorithm. This technique is quite good and covers all the things needed for vehicle detection and identification system. But main drawback of this system is that it concentrates on feature extraction and labeling of vehicle mages and not on prediction using machine learning where machine learning improves accuracy of the system.

In 2016 Suganya.R. et al. [4] presented the paper focusses mainly on lesion detection and classification of dermoscopy images detection using K-Means clustering and SVM algorithm. This technique is quite good and covers all the things needed for lesion detection and classification system. But main drawback of this system is that it concentrates on K-Means algorithm for segmentation and feature extraction, whereas the SIFT algorithm that we propose is far more accurate and return the features for noisy images with greater efficiency thus improving and helping the machine learning algorithm with better accuracy.

In 2020 Vidya M. et al. [5] presented the paper focusses mainly on skin lesion detection using machine learning technique. This technique is quite good and covers all the things needed for early skin lesion detection system. But main drawback of this system is that it concentrates on GLCM algorithm for segmentation and feature extraction, whereas the SIFT algorithm that we propose is far more accurate and return the features for noisy images with greater efficiency thus improving and helping the machine learning algorithm with better accuracy.

In 2019 Pham et al. [6] presented the paper focusses mainly on skin cancer detection using images. This technique is quite good and covers all the things needed for skin cancer detection system. But main drawback of this system is that it concentrates on HSV algorithm for segmentation and feature extraction, whereas the SIFT algorithm that we propose is far more accurate and return the features for noisy images with greater efficiency thus improving and helping the machine learning algorithm with better accuracy.

In 2020 krishna et al. [7] presented the paper focusses mainly on mainly on lung cancer detection using SIFT and Bag of Words algorithm. This technique is quite good and covers all the things needed for detecting lung cancer. But main drawback of this system is that it concentrates on lung cancer and does not define how it can be used for other diseases such as skin cancer. It uses only one algorithm, where as we propose to use SVM and Naïve Bayes to check and get better accuracy.

In 2020 Rao et al. [8] presented the paper focusses mainly on mainly on Bone fracture detection using SIFT and Bag of Words algorithm. This technique is quite good and covers all the things needed for detecting bone fracture. But main drawback of this system is that it concentrates on bone fracture and does not define how it can be used for other diseases such as skin cancer. It uses only one algorithm, where as we propose to use SVM and Naïve Bayes to check and get better accuracy.

III.METHODOLOGY

This section will study the mathematical conditions and algorithms to be used for designing a automatic skin cancer diagnosis framework. These are explained as follows:

A. Mathematical Model

Our skin cancer diagnosis technique can be explained in two sets with probability, success and failure conditions.

1) Feature extraction Module

Set (F)= {F0, F1, F2, F3, F4, F5}
F0∈F = Get skin lesion database.
F1∈F = Apply SP-SIFT on image database.
F2∈F = Extract features from images.
F3∈F = Establish communication with cloud.
F4∈F = Upload extracted feature to cloud.
F5∈F = View results.

2) Machine learning Module

Set (M)= {F3, M0, M1, M2, M3, F5}
F0∈M = Get skin lesion database.
M0∈M = Download extracted features from cloud.
M1∈M = Create training dataset from features.
M2∈M = Pass a test image and create testing dataset.
M3∈m = Apply SVM and Naïve Bayes to classify test image
And return result into classes as melanoma, non-melanoma and benign.
F5∈M = View results.

So, by studying the sets we come to notice that many elements are common in both modules and used in coordination in both sets so they be placed as

$$x \in F \cap M \text{ if } x \in F \text{ and } x \in M$$

Thus, the probability of intersection of elements in both modules can be given as

$$P(F \cap M) = P(F)+P(M)$$

So, intersection of common elements can be shown as

$$F \cap M = \{F3, F5\}$$

The conditional probability of both modules using the same element can be shown as

$$P(F \mid M) = \frac{P(F \cap M)}{P(M)}$$

Thus, we conclude that our malicious skin cancer diagnosis framework's success and failure will depend upon the quality of image database, i.e., if the quality of image database is not good the SP-SIFT feature extraction and SVM and Naïve Bayes classification results will not be proper, thus this is a case of failure, so our framework supports NP-Hard and not NP-Complete.

B. Algorithms Used

Our skin disease classification process can be explained using following algorithmic steps.

```
procedure CLASSIFY
    skinImages[]=Fetch skin cancer image database
    imageCount=skinImages.length()
    if imageCount == 0 then
        for i = 0;i < imageCount;i++ do
            resizeStatus=resizeImages(skinImages(i))
            if resizeStatus == True then
                Extract feature Keypoints using SIFT
                Create training dataset
            else
                result ←Exit procedure
            end if
        end for
        fetchStatus==Fetch images sent by user
        if fetchStatus == True then
            resizeStatus=resizeImages(fetchImages(rowNo))
            if resizeStatus == True then
                Extract feature Keypoints using SIFT
                Create testing dataset
                Train and apply SVM and Naive Bayes
                View and Send Classification result to the user.
            else
                result ←Exit procedure
            end if
        else
            result ←Exit procedure
        end if
    else
        Repeat Step 2:
    end if
    View movie review classification.
end procedure
```

IV. PROPOSED SYSTEM

This section will study the mathematical conditions and algorithms to be used for designing an automatic skin cancer diagnosis framework. These are explained as follows:

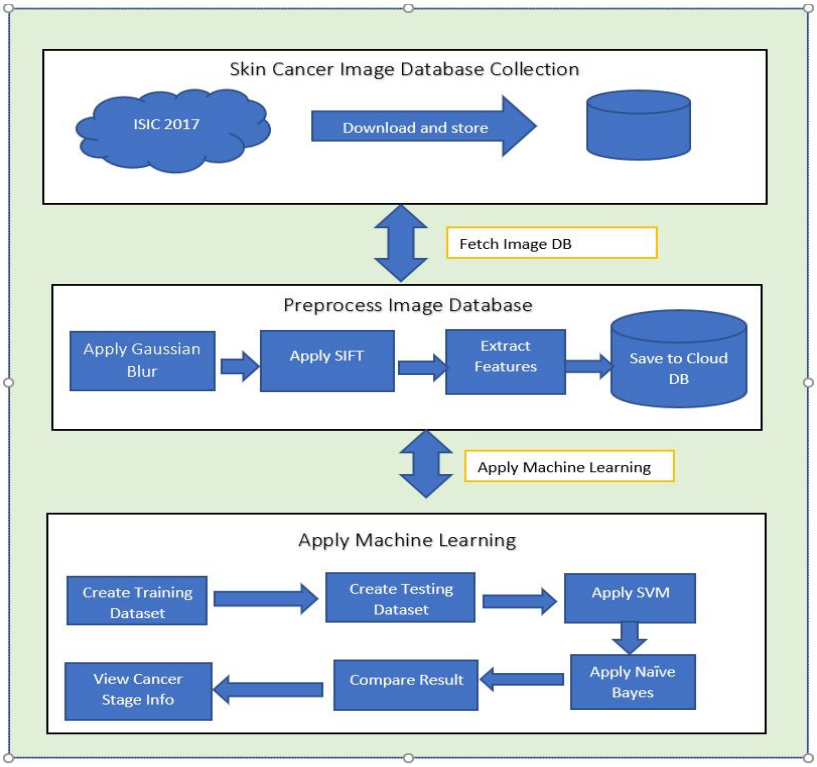


Fig.2. System Architecture Diagram.

This This section is mainly divided in 3 main modules with other sub parts in them. The text that follows explains the modules with a block diagram or system architecture as shown in Fig.2. to illustrate them. The working of the framework is explained as:

A. Skin Lesion Image Database

This module will be proposed to accumulate image and create an image DB. We are going to use ISIC 2017 skin cancer database as a reference images in our project. We are going to download the set of images which are defined in three stages melanoma, non-melanoma and benign. The images will be downloaded and saved in three different folders named melanoma, non-melanoma and benign on the local drive as shown in system architecture. The images will be categorized in to three cancer stages as they will be followed in the next project stages.

B. Preprocess Skin lesion image DB:

This module will also be proposed to pre-process image DB for machine learning. The images stored on local drive will be accessed to perform pre-processing steps. Then SP-SIFT algorithm will be applied on the images and features will be extracted. The extracted features will give values for parameters such as contrast, correlation etc. The extracted features will then be categorized in to three types melanoma, non-melanoma and benign. The grouped features will then be stored in cloud DB using google drive sub cloud google spreadsheet for later use remotely. Thus, it can be accessed to create training dataset which will be used for machine learning. The main aim of pre-processing step is to get image database ready for machine learning.

C. Apply Machine Learning

This This module will be proposed to find the stage of a skin cancer in a test image. It will use the features that are stored in cloud DB. First a training dataset is created which will be used to train machine learning algorithms. The training dataset will have three classes melanoma, non-melanoma and benign. Then a test image will be passed to get stage of a cancer in that image. The test image will follow the pre-processing steps and create a testing dataset from it with same three classes as training dataset. Then both training and testing datasets will be given to Naïve Bayes algorithm which will predict the class in which the test image resides i.e., whether the cancer is melanoma, non-melanoma or benign. Then same training and testing datasets will be given to SVM algorithm which will predict the class in which the test image resides i.e., whether the cancer is melanoma, non-melanoma or benign. Both predictions than will be compared for the accuracy of the result.

V. RESULTS AND DISCUSSION

Thus, to explain the above proposed system we have created 2 applications. One for admin desktop where he can analyze the Skim images sent to him from the user's app and he can send the skin classification results back to the user using cloud.

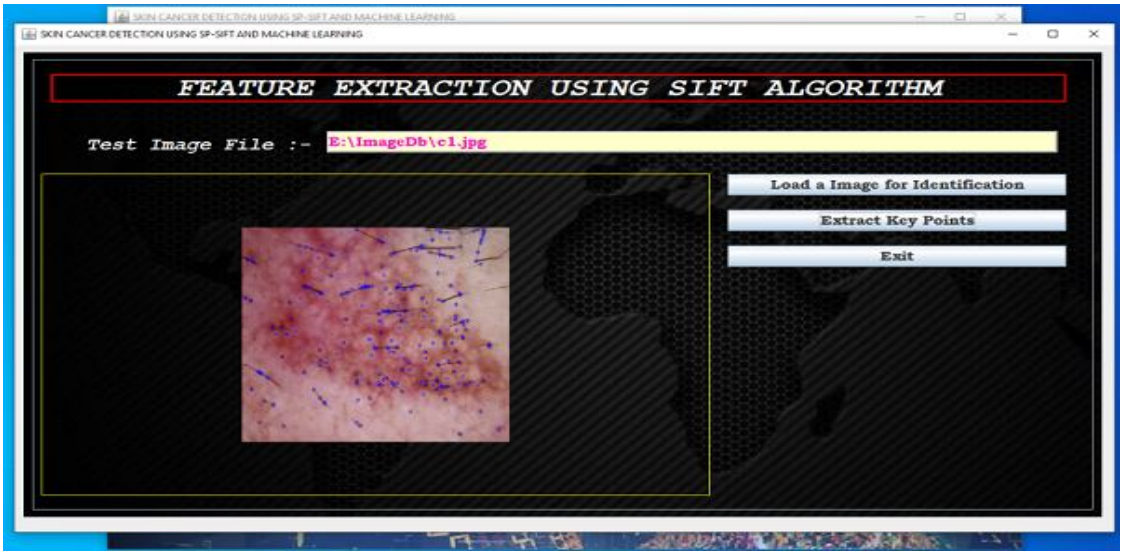


Fig.3 Demonstration of working of SIFT.

In Fig.3 it shows how SIFT algorithm works and how it finds the key points. These key points are used to create a training dataset after resizing all the images to a same size and then extracting features and creating a training dataset of these features with 2 classes benign and malignant.

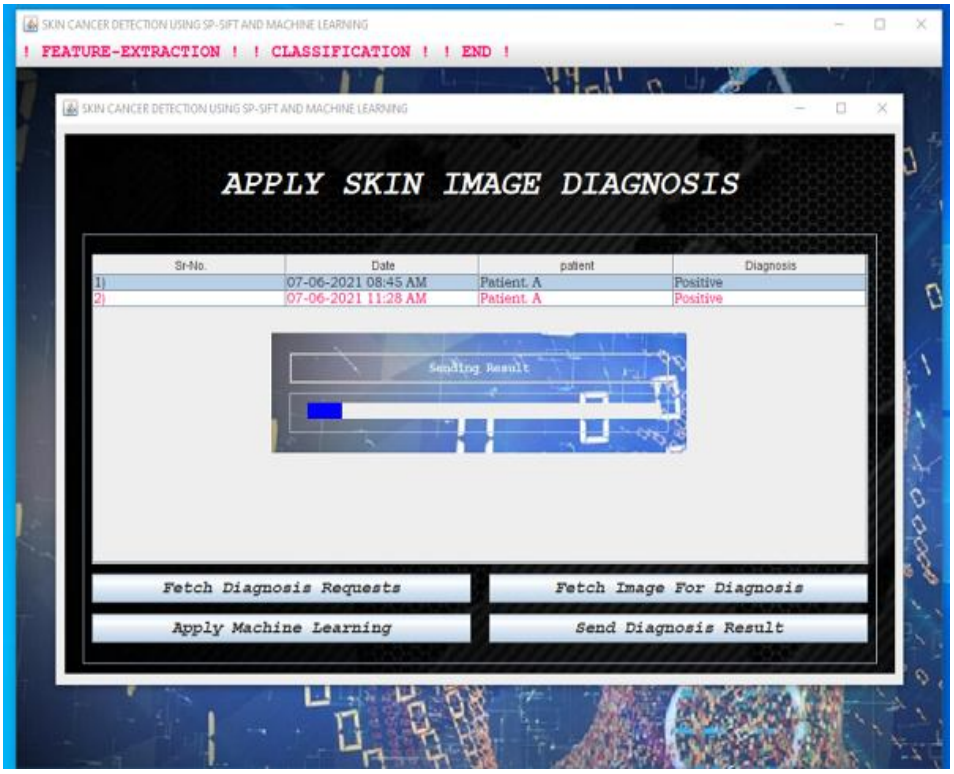


Fig.4. Sending skin diagnosis results.

In Fig.3 it shows how SIFT algorithm works and how it finds the key points. These key points are used to create a training dataset after resizing all the images to a same size and then extracting features and creating a training dataset of these features with 2 classes benign and malignant.

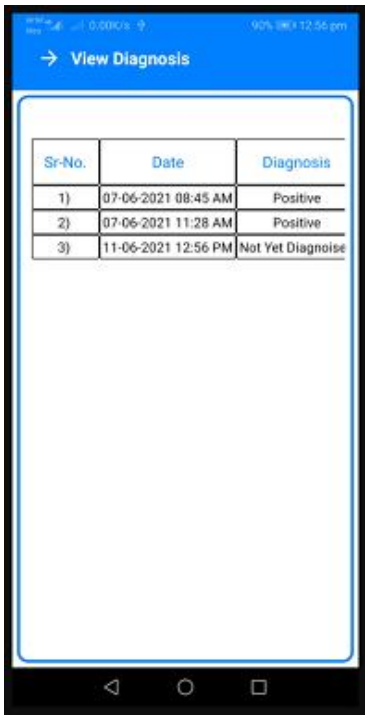


Fig.5. Skin image diagnosis results.

In Fig.5 shows how the results are viewed on the user side after he sends an image.

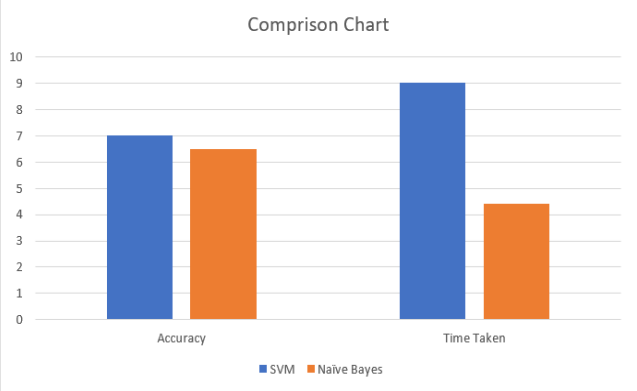


Fig.6. Skin image diagnosis results.

Thus, the results in Fig.6. suggests that SVM has more accuracy in skin cancer diagnosis than Naïve Bayes. But it also shows that the time to give result or process to complete skin cancer diagnosis is more for SVM than Naïve Bayes. The accuracy of SVM is more than naïve bayes as we have provided a good number of features in the dataset so SVM gets trained well and returns more accurate than naïve bayes. If we pass fewer features in the dataset the accuracy of naïve bayes is more than that of SVM.

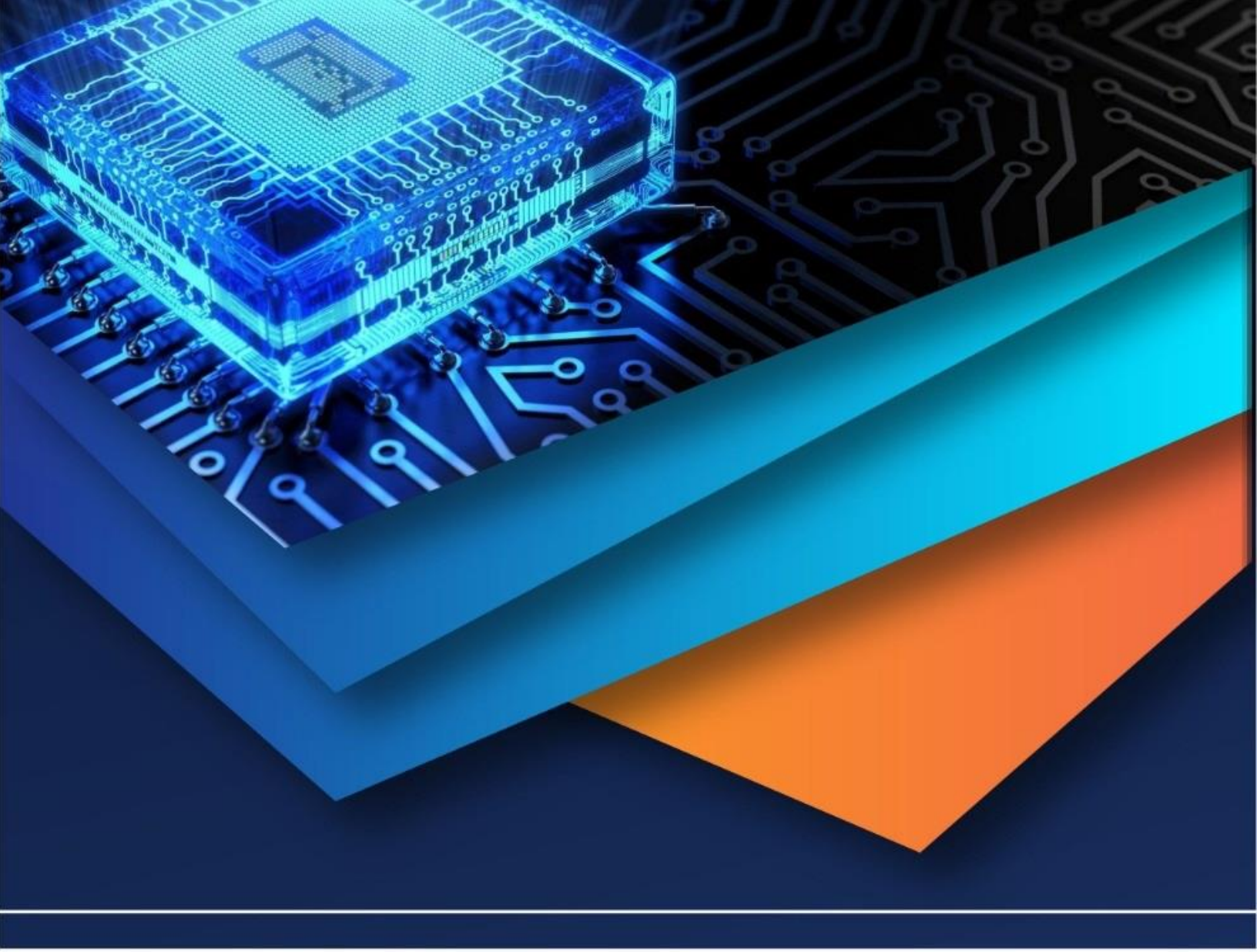
VI. CONCLUSIONS

In this paper we conclude and discussed to develop a novel approach of skin cancer diagnosis system using image processing, machine learning algorithm and cloud together. We have compared our feature extraction algorithmn SIFT with other algorithms mentioned in [4][5][6] like K-Means, GLCM and HSV and illustrated why we are using the algorithm. We are making use of cloud technology to store image features so that they can be accessed from anywhere any time. We are concluding to try to combine and make use of Java and Python or any other languages together to get effective output from the frame work. We conclude to generate good training dataets for SVM and Naïve Bayes algorithms so to get more accurate results. Thus we conclude that our framework will be helpful in preventing unneccsary deaths that are caused due to late diagnosis of skin cancer and thus help a medical practioner to concentrate on treatement rather than diagnosis.

REFERENCES

[1] Fulgencio Navarro, Marcos Escudero-Viñolo and Jesús Bescós., - Accurate segmentation and registration of skin lesion images to evaluate lesion change., IEEE-2019.
[2] Feng Guo, Jie Yang, Yilei Chen, Bao Yao., - Research on Image Detection and Matching Based on SIFT Features., IEEE-2018.
[3] Tu Yawen and Guo Jinxu., - Research on Vehicle Detection Technology Based on SIFT Feature., IEEE-2018.
[4] Suganya R, - An Automated Computer Aided Diagnosis of Skin Lesions Detection and Classification for Dermoscopy Images, in IEEE-2016.
[5] Vidya M and Dr. Maya V Karki, - Skin Cancer Detection using Machine Learning Techniques, in IEEE-2020.
[6] Tri Cong Pham, Giang Son Tran, Thi Phuong Nghiem, Antoine Doucet, Chi Mai Luong and Van-Dung Hoang, - AComparative Study for Classification of Skin Cancer, in IEEE-2019.
[7] Azmira Krishna, P.C.Srinivasa Rao and CMAK Zeelan Basha, - Efficient Computerized Lung Cancer Detection Using Bag of Words, in IEEE 2020.
[8] L.Jagjeevan Rao, P.Neelakanteswar, Madupu Ramkumar and Azmira Krishna, - An Effective Bone Fracture Detection using Bag-Of- Visual-Words with the features extracted from SIFT, in IEEE-2020.
[9] H. Cheng, - Computer-aided detection and classification of microcalcifications in mammograms: a survey, Pattern recognition, 36(12), 2967-2991, 2003.
[10] M. E. Celebi, - Lesion border detection in dermoscopy images, Comput. Med. Imag. Graph., 33(2), 148-153, 2009.
[11] C.A.Z. Barcelos, V.B. Pires., - An automatic based nonlinear diffusion equations scheme for skin lesion segmentation, Applied Mathematics and Computation, 215, 251–261, 2009.
[12] Jaisakthi, Seetharani Murugaiyan, Palaniappan Mirunalini, and Chandrabose Aravindan, -Automated skin lesion segmentation of dermoscopic images using Grab Cut and k-means algorithms, IET Computer Vision 12, no. 8 (2018): 1088-1095.
[13] Chung, Do Hyun, and Guillermo Sapiro, -Segmenting skin lesions with partial-differential-equations-based image processing algorithms, IEEE transactions on Medical Imaging 19, no. 7 (2000): 763-767.
[14] Hemalatha RJ, Thamizhvani TR, Dhivya AJ, Joseph JE, Babu B, Chandrasekaran R, -Active Contour Based Segmentation Techniques for Medical Image Analysis. Medical and Biological Image Analysis, 2018 Jul 4:17.
[15] SALIH, SHAKAR H., and S. H. E. R. E. E. N. AL-RAHEYM, -Comparison of Skin Lesion Image Between Segmentation Algorithms, Journal of Theoretical and Applied Information Technology 96, no. 18 (2018).
[16] Li, Yuexiang, and Linlin Shen, -Skin lesion analysis towards melanoma detection using deep learning network, Sensors 18, no. 2 (2018): 556.
[17] Kasmi, R. and Mokrani, K., -Classification of malignant melanoma and benign skin lesions: implementation of automatic ABCD rule. IET Image Processing, 10(6), pp.448-455.

[18] Goel, Rashi, and Saranjeet Singh, -Skin cancer detection using glcm matrix analysis and back propagation neural network classifier, International Journal of Computer Applications 112, no. 9 (2015).
[19] Bakheet, Samy, -An svm framework for malignant melanoma detection based on optimized hog features, Computation 5, no. 1 (2017): 4. Khan, Muhammad Qasim, Ayyaz Hussain, Saeed Ur Rehman, Umair
[20] Khan, Muazzam Maqsood, Kashif Mehmood, and Muazzam A. Khan, -Classification of Melanoma and Nevus in Digital Images for Diagnosis of Skin Cancer, IEEE Access 7 (2019): 90132-90144.
[21] Victor, Akila, and M. Ghalib, -Automatic detection and classification of skin cancer, International Journal of Intelligent Engineering and Systems 10, no. 3 (2017): 444-451.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY