2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN)

Classification of Skin diseases using Image processing and SVM

N Vikranth Kumar
VIT University
Vellore, India
vickievikranth@gmail.com

P Vijeeth Kumar
VIT University
Vellore, India
vijeeth.vj8@gmail.com

K Pramodh
VIT University
Vellore, India
Pramodh258@gmail.com

Prof. Yepuganti Karuna
VIT University
Vellore,India
karun@vit.ac.in

***Abstract*— Skin diseases such as Melanoma and Carcinoma are often quite hard to detect at an early stage and it is even harder to classify them separately. Recently, it is well known that, the most dangerous form of skin cancer among the other types of skin cancer is melanoma because it is much more likely to spread to other parts of the body if not diagnosed and treated early. In order to classify these skin diseases, "Support Vector Machine (SVM)" a Machine Learning Algorithm can be used. In this paper, we propose a method to identify whether a given sample is affected with Melanoma or not. The steps involved in this study are collecting labelled data of images that are pre-processed, flattening those images and getting the pixel intensities of images into an array, appending all such arrays into a database, training the SVM with labelled data using a suitable kernel, and using the trained data to classify the samples successfully. The results show that the achieved accuracy of classification is about 90%.**

Keywords—Melanoma, SVM, array, images, Carcinoma, classify, machine, algorithm

I. INTRODUCTION

A. *Background and Motivation*

Skin diseases are one of those set of diseases whose number has been largely increasing day by day. Only in India, about 200 million people suffer from one or the other forms of skin diseases. People often neglect skin diseases and do not take necessary treatment. This is especially seen in rural and economically backward areas due to many factors such as lack of awareness, poverty and lack of resources etc. this is even higher when it comes to the case of Melanoma skin cancer. It is reportedly found that about 132,000 melanoma skin cancers occur globally each year [1]. When the people tend to approach a physician, it is quite difficult for the physician in order to exactly detect the type of skin diseases the patient is getting affected with. Especially when it comes to the diseases like Melanoma, it is quite hard to differentiate without any tests being conducted. In men, it is often found on the skin on the head, on the neck, or between the shoulders and the hips while, in women, it is often found on the skin on the lower legs or between the shoulders and the hips [2]. Besides SVM, another technique can also be used to classify among diseases. That is the classification using "Neural Networks" [3]. However, SVM is a better technique to classify than Neural Networks because they have a strong founding theory. SVMs reach the global optimum due to quadratic programming, they have no issue for choosing a proper number of parameters, Also, SVMs are less prone to over fitting and they need less memory to store the predictive model also yielding results that are more readable.

B. *Contribution*

In the literature, we found that many authors working using different algorithms to identify diseases but there is very limited research on using one particular method to classify two or more different diseases. Here, we proposed an efficient technique in which the database of pre-processed images are trained and tested and are classified using SVM, a machine learning based algorithm to identify whether the skin lesion is benign or malignant. This will be very helpful in the diagnosis of the Melanoma skin cancer efficiently.

C. *Related Work*

In order to detect Melanoma various authors have done in the field of image processing. Some have used MATLAB software to analyze and investigate the best formats to carry out the analysis. Pre-processing techniques such as Hair removal, centering the image, shading effect, vignette and black-border cropping [4]. For segmentation, techniques such as Otsu's Thresholding [5], color space transformation, Watershed algorithm, c-means algorithm were used [6]. Works such as introducing the Image based screening techniques to differentiate similar diseases and Multi-SVM classifiers were done in this field [7].

978-1-5386-9353-7/19/$31.00 ©2019 IEEE

II. METHODOLOGY

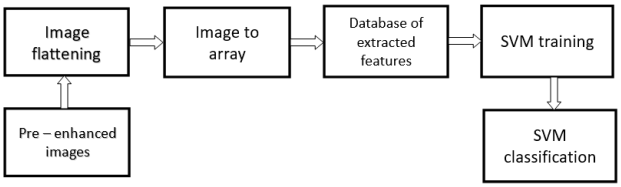


Figure 1: The proposed Block Diagram

A. *Pre-enhanced images*

The statistical data is taken from an image database [8] that is verified to be prone to a particular disease. This database of images are pre-enhanced i.e.; they are undergone techniques such as hair removal, centering of the image, softening.

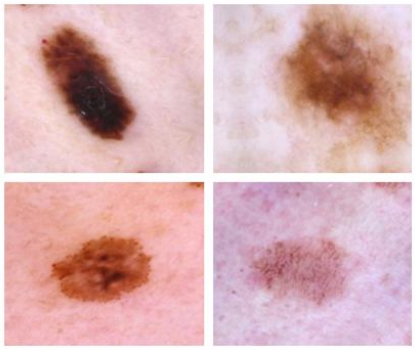


Figure 2: Pre enhanced images taken from verified website

B. *Image flattening and image to array*

The pre-enhanced images are converted into the RGB format and all the pixel values of the image in RGB format are converted into a 1-Dimensional array. The recorded pixel intensities are scaled in between (0, 1). Here, we took a 64*64-pixel format. So around 12288 values for each image have been recorded.

C. *Database of image Intensities*

All the 1-Dimensional arrays are arranged into the form of a database by appending all the image pixel intensities of all the images. By doing this, a database is created with all the pixel intensities of all the images.

As it is observed in figure 3, all the pixel intensities are taken and are formed into a database, which is further then used to train the SVM as it is given as the input data.

	A	B	C	D	E	F	G	H	I
1	0.792157	0.639706	0.682353	0.786029	0.626716	0.643627	0.781863	0.615196	0.618627
2	0.889449	0.566345	0.621864	0.886248	0.560758	0.596542	0.8937	0.54938	0.595473
3	0.857598	0.794853	0.908578	0.82598	0.763235	0.876961	0.81152	0.748775	0.8625
4	0.630465	0.481752	0.565552	0.61929	0.503669	0.560815	0.642823	0.509815	0.577562
5	0.691131	0.595244	0.5964	0.702934	0.606154	0.647296	0.710352	0.607127	0.642896
6	0.146527	0.074372	0.082901	0.121113	0.060574	0.068907	0.114262	0.066288	0.075042
7	0.770588	0.707843	0.813725	0.748039	0.685294	0.791176	0.721324	0.646078	0.755882
8	0.589951	0.580258	0.637416	0.63974	0.58673	0.688366	0.650019	0.639078	0.720661
9	0.825	0.713725	0.804657	0.82451	0.710784	0.797059	0.820098	0.694608	0.776961
10	0.911129	0.554389	0.575383	0.92361	0.570247	0.59257	0.90306	0.55422	0.561156
11	0.910658	0.665097	0.739005	0.911439	0.668302	0.754086	0.905825	0.654956	0.749131
12	0.926949	0.591261	0.573886	0.94277	0.591517	0.582564	0.940805	0.592165	0.56906
13	0	0	0	0.00175	0.00175	0.00175	0.000306	0.000306	0.000249
14	0.925429	0.616337	0.727868	0.927505	0.594443	0.723713	0.921404	0.611765	0.726042
15	0.828431	0.75	0.855882	0.81348	0.726225	0.835049	0.807843	0.711275	0.824265
16	0.777696	0.532598	0.53799	0.803431	0.557353	0.57451	0.783333	0.546324	0.551225
17	0.541004	0.336837	0.384559	0.544941	0.343313	0.422461	0.564648	0.380817	0.459367
18	0	0	0	0	0	0	0	0	0
19	0.657414	0.548418	0.607487	0.650751	0.544799	0.615859	0.658766	0.544861	0.631675
20	0.675119	0.635061	0.669593	0.67981	0.638699	0.67014	0.682219	0.639036	0.670523
21	0.545872	0.440782	0.457403	0.554033	0.464959	0.451314	0.565081	0.463109	0.459199

Figure 3: Database of pixel intensities of all the images

D. *Training the SVM*

A Support Vector Machine is nothing but a machine-learning algorithm, which can classify among two or more classes [9], [10], and [17]. The classification happens because of different kernels, which are used as hyper planes to differentiate among the classes [15]. The accuracy and precision of the SVM mainly depends on the Kernel used and the boundary values defined [4]. Therefore, a suitable Kernel has to be taken in order to achieve better results. In this study, to differentiate between Melanoma and Non-Melanoma, a "Linear kernel" is used. Linear kernels are used when the classes that are to be separated do not have many features in common [11]. Linear kernel is also one of the simplest of all the kernels available. When we want to classify two classes, which are having more features in common, then other kernels such as Polynomial Kernels are to be used to achieve better accuracy and precision [12]. Gamma kernel is used to define the boundary values of the SVM [12]. Linear kernel's equation is as follows:

$$f(x,y) = x^T y + c$$

This equation involves calculating the inner products of a new input vector (x) with all support vectors in training data [13]. The coefficients 'y' is the distance from the hyper plane to the feature and 'c' is an optional constant [13]. All the values, which were added into a database as mentioned, are given as input to the SVM and is trained to differentiate the classes as it is labeled data.

E. *SVM classification*

When an unknown data is given to the SVM, it classifies the sample based on the training samples [10], [16]. Hence, SVM classifies the image whether it belongs to Melanoma or Non- Melanoma. As the "enum" function

is used, the algorithm identifies the matched Melanoma samples as 0 and unmatched samples as 1.

III. Simulation Results

When working with SVM, The result depends on how well the SVM is being trained. So, more the number of images the SVM is trained with, the higher is the accuracy. We tried training the SVM with different set of images as input and we achieved the accuracy as follows:

Images(Training)	Precision	Time(min)
600	70%	9
900	80%	13
1400	84%	37
1700	90%	46

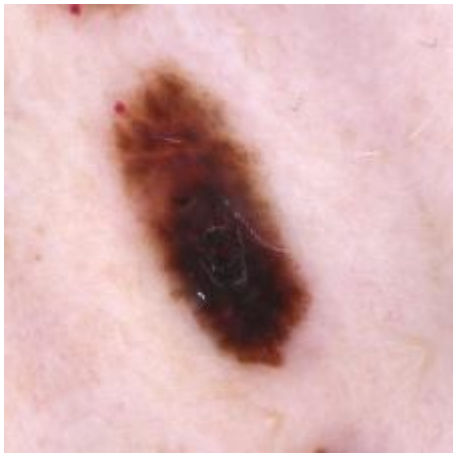
Table 1: Table showing number of images, precision achieved and time taken for the code to be compiled

As we can observe, with the increase in the number of images to be trained, the precision is increased also with an increase in the compilation time. In python about 600 images data were approximately compiled in 9 min whereas in MATLAB, it took about 2 hours. This is the main reason for continuing this study in python.

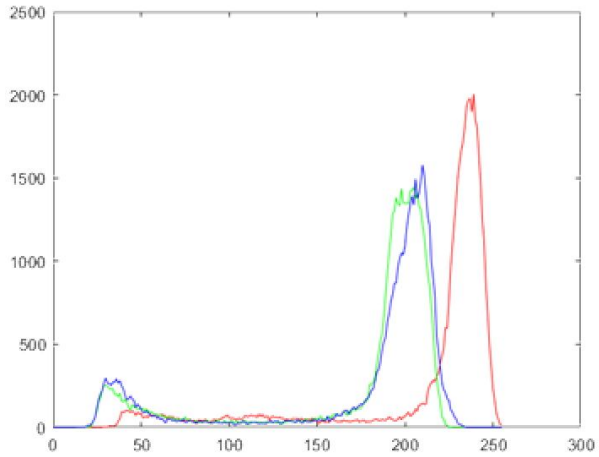
Image no.	Predicted result
1	0
2	1
3	0
4	1
5	0
6	1
7	0
8	1

Table 2: Classification of Melanoma and Non-Melanoma

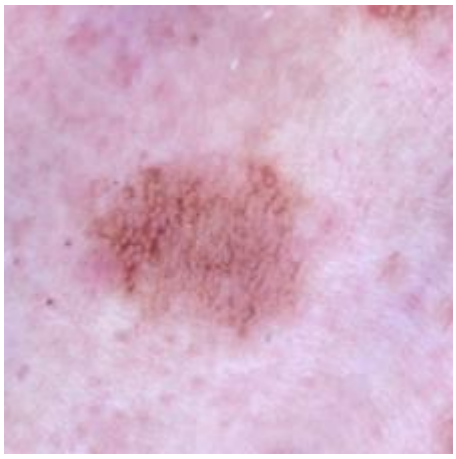
Table 2 shows the classification done by SVM as it identifies the samples belonging to Melanoma and Non Melanoma. Here '0' represents Melanoma and '1' represents Non Melanoma.



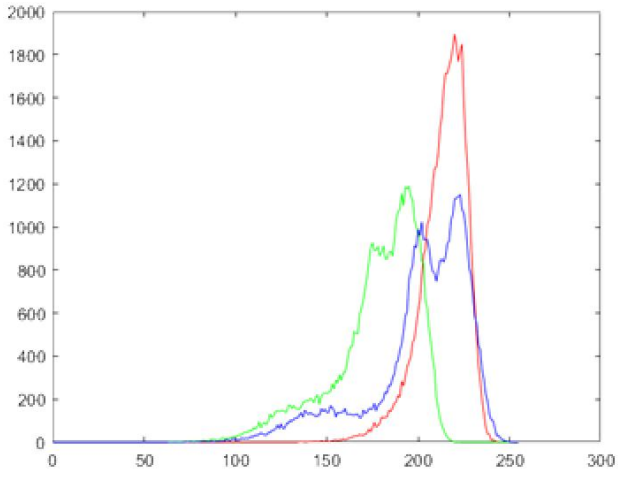
(a)



(b)



(c)

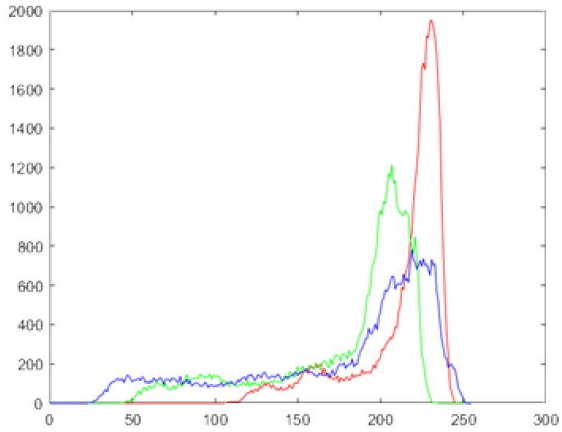


(d)

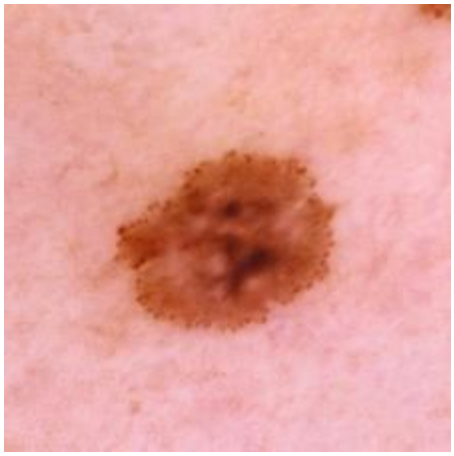
Figure 4: Image histograms of matched results

In figure 4, (a) & (c) represent Melanoma images and (b) & (d) represent the respective image histograms of matched results. It can be clearly observed that the RGB planes of both the images are very similar to each other thus proving that the images are malignant prone images.

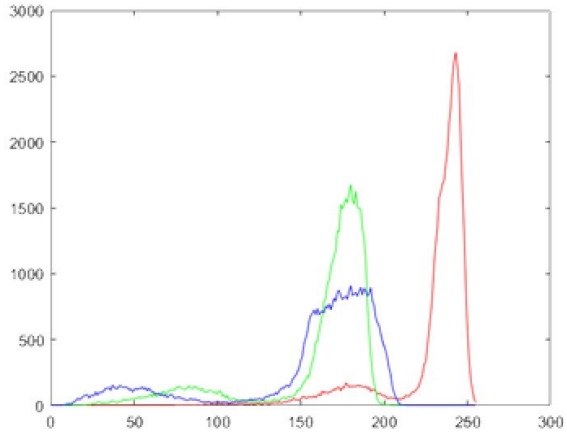
(a)



(b)



(c)



(d)

Figure 5: Image histograms of unmatched results

In figure 5, (a) represents Melanoma and (c) represents Non-Melanoma image and (b) & (d) represent the respective image histograms of unmatched results.

IV. Conclusion and Future Work

In this work, we presented that SVM can be effectively used to classify among the samples containing Melanoma and Non Melanoma. It is observed that better results and precision can be achieved when the SVM is trained with more number of images. The future work can be developing a product that can differentiate between two different types of cancers such as Melanoma and Carcinoma. Higher accuracy can be attained when other kernels such as Polynomial Kernels are used.

V. References

[1] Skin cancers, World Health Organization, Available: [online] http://www.who.int/uv/faq/skincancer/en/index1.html Accessed December, 2018.

[2] Landis, Sarah H., et al. “Cancer Statistics, 1999.” CA: A Cancer Journal for Clinicians 49.1(1999):8-31

[3] Codella, Noel, et al. "Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images." *International workshop on machine learning in medical imaging*. Springer, Cham, 2015.

[4] Mustafa, Suleiman, Ali Baba Dauda, and Mohammed Dauda. "Image processing and SVM classification for melanoma detection." *2017 International Conference on Computing Networking and Informatics (ICCNI)*. IEEE, 2017.

[5] Vala, Hetal J., and Astha Baxi. "A review on Otsu image segmentation algorithm." *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*2.2 (2013): 387-389.

[6] Manerkar, Mugdha S., et al. "Automated skin disease segmentation and classification using multi-class SVM classifier." (2016).

[7] Manerkar, Mugdha S., et al. "CLASSIFICATION OF SKIN DISEASE USING MULTI SVM CLASSIFIER." *3rd International Conference on Electrical, Electronics, Engineering Trends, Communication, Optimization and Sciences—2016*.

[8] Image dataset: https://www.kaggle.com/drscarlat/melanoma

[9] Bono, Aldo, et al. “The ABCD system of melanoma detection.” Cancer 85.1 (1999): 72-77.

[10] Alquran, Hiam, et al. "The melanoma skin cancer detection and classification using support vector machine." *2017 IEEE Jordan*

Conference on Applied Electrical Engineering and Computing Technologies (AEECT). IEEE, 2017.

[11] Rosipal, Roman, Leonard J. Trejo, and Bryan Matthews. "Kernel PLS-SVC for linear and nonlinear classification." *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*. 2003.

[12] Smits, Guido F., and Elizabeth M. Jordaan. "Improved SVM regression using mixtures of kernels." *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290)*. Vol. 3. IEEE, 2002.

[13] Cesar Souza (2010,March 17) "http://crsouza.com/2010/03/17/kernel-functions-for-machine-learning-applications

[14] Fatima, Ruksar, Mohammed Zafar Ali Khan, and K. P. Dhruve. "Computer aided multi-parameter extraction system to aid early detection of skin cancer melanoma." *International Journal of Computer Science and Network Security* 12.10 (2012): 74-86.

[15] Lv, Xiao. "A novel defect inspection approach using image processing and support vector machines in bolts." *2015 Seventh International Conference on Measuring Technology and Mechatronics Automation*. IEEE, 2015.

[16] Abe, Shigeo. *Support vector machines for pattern classification*. Vol. 2. London: Springer, 2005.

[17] Dreiseitl, Stephan, et al. "A comparison of machine learning methods for the diagnosis of pigmented skin lesions." *Journal of biomedical informatics* 34.1 (2001): 28-36.