ANALYSIS AND CLASSIFICATION OF LEUCODERMA USING SUPPORT VECTOR MACHINE AND K-NEAREST NEIGHBOUR ALGORITHM

Anjali Mittal[1], Harish Chandra Maurya[2]

[1] Research Scholar, Computer Science and Engineering, Bhagwant University Ajmer, Rajasthan, India
[2] Assistant Professor, Computer Science and Engineering, Bhagwant University Ajmer, Rajasthan, India

ABSTRACT

Support Vector Machines (SVM) as well as k- Nearest Neighbour Algorithm (KNN) is very powerful tools which can be utilized for skin disease recognition. They are being used in a large array of different areas including medicine. This thesis exemplifies the applicability of computer science in medicine especially in skin disease detection. Leucoderma or commonly known as Vitiligo is a very common disease among all other skin diseases. In the research work I am comparing SVM classifier and KNN, two the most effective techniques for data classification (especially for medical diagnoses), implemented using matlab tool and classify the patients affected by Leucoderma. Experimental results show that the proposed methodology efficiently discriminates between a leucoderma skin and a normal skin.

Keyword: - *Dermatology, Support vector machine, leucoderma.*

1. INTRODUCTION

Skin diseases are now very common all over the world. Different kinds of allergies are also becoming more common. Many of these diseases are very dangerous, particularly when not treated at an early stage. Dermatologists have at their disposal large catalogues with pictures of skin segments, which they use as a reference for diagnosis of their patients cases' skin ailments. However it may be difficult even for experienced doctors, to make correct diagnosis because many symptoms look very similar to each other, even though they are caused by different diseases. All details such as colour, size or density of the skin changes are important. Modern medicine is looking for solutions, which could help doctors with any aspect of their work using the new technology. Such tools already exist, to our knowledge they concentrate on analysis of the colour of skin changes and UV photography. The goal of this project is to make a system to recognize skin disease Leucoderma or commonly known as Vitiligo using Support Vector Machines (SVM) and k-Nearest Neighbour Algorithm. After testing both methods results will be compared. System should learn from the set of skin segments images taken by digital camera. After that it should return the probability of existence of any recognized disease, based on the same type of image made by user. Images should have the same size and should be taken from the same distance. Some practical study may display need of using other information (not only pictures) to train the network, for example part of the body where the symptoms were found or if the patient feels pain or tickle

2. RELATED WORK

A lot of work has been done in the field of leather in recent year's disease detection using computer applications. Most the implementation part was done using Support Vector Machine. We studied these researches, some of which we want to discuss in this document.

2.1 Automatic Detection of Ringworm:

Srimanta Kundu, Nibaran Das and Mita Nasipuri, used Svm for the automatic detection of ringworm in their paper using For the Support Vector Machine classifier; an open source software LibSVM tool is used. In general, a classification task usually involves with training and testing data which consist of some data instances. Each instance in the training set contains one "target value" (class labels) and several "attributes" (features). The goal of SVM is to produce a model which predicts target value of data instances in the testing set which are given only the attributes. Before considering the data directly from the linearly scaling each attribute to the range [-1, +1] or [0, 1].
Given a training set of instance-label pairs (x_i, y_i);

$$i = 1 \ldots N;$$

Where $x_i \in R^n$ and $y \in \{1, -1\}$, the support vector machines (SVM) require the solution of the following optimization problem:

$$\min (w, w_0, €)\ \tfrac{1}{2} w^T w + C$$
$$\text{Subject to } y_i (w^T x_i + b) \geq 1 - €_i,\ €_i \geq 0$$

It's not difficult to generalize this linear program to the nonlinear case replacing x_i with a nonlinear function $M(x_i)$:

$$\min (w, w_0, €)\ \tfrac{1}{2} w^T w + C$$
$$\text{Subject to } y_i (w^T M(x_i) + b) \geq 1 - €_i,\ €_i \geq 0$$

Fig 1: sample of Ringworm images

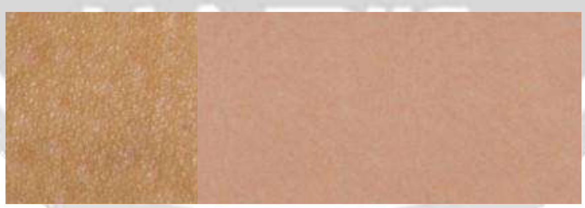

Fig 2: sample of normal skins

According to their research they achieved 90% success rate in the automatic detection of ringworm using support vector machine.

2.2 Skin Cancer

Skin diseases are now very common all over the world. For example the number of people with skin cancer has doubled in the past 15 years. "Currently, between 2 and 3 million non-melanoma skin cancers and 132,000 melanoma skin cancers occurs globally each year.Svm also applied to some medical tools. One of them, proposed in [33], identifies patients with breast cancer for whom chemotherapy could prolong survival time. In the experiment SVM was used for classifying patients into one of three possible prognostic groups: Good, Poor or Intermediate. After classification of 253 patients, 82.7% test set correctness was achieved. Another application of SVM in medicine is described in [18]. A system for cancer diagnosis used the DNA micro-array data as a classification data set. Diagnosis error obtained by above mentioned system was smaller than in systems which uses other known methods. Reduction of the error achieved 36%. SVM is relatively new method of classification and it expands very quickly. That will certainly cause wider use of SVM in different areas, also in medicine.

2.3 Skin Mole localization using DoG filters and a support vector Machine

Mole pattern changes are important cues in detecting early signs of melanoma, a deadly skin cancer [4]. Early detection is especially important for melanoma because, while advanced cases are not curable, the disease can be cured if detected early [16]. However, a principled system to register mole pattern changes is currently lacking. In fact, a major burden on the dermatological workforce is in manual surveillance of pigmented lesions, which is both time consuming and prone to human error. The block diagram representation of the mole localization stage is shown in Figure 8. In this section, DoG scale-space filters and the designed support vector machine (SVM) classifier will be introduced. Since the size of moles can vary, moles should be searched in a multi-scale fashion. The DoG filter is applied to RGB color channels separately, and the set union of the output maxima over scale in each channel are considered possible mole candidates. When combining the DoG maxima, any mole candidates' occurring within a radius of another mole candidate is eliminated. Once mole candidates are localized, regions around mole candidates are cropped from the *hair-removed* image. The width of mole candidate patches is 2 £ *p* 2 times the radius of each DoG maximum. Cropped mole candidates are classified as moles using a support vector machine (SVM) classifier. An SVM is a powerful tool to both generalize and classify objects: LIBSVM [1] was used to build the SVM classifier

3. PRESENT WORK

A large database of leucoderma affected images has been used to test the classifiers. The performance of the classifiers varies with original color of the skin. for example it is a little difficult to indentify vitiligo for an extremer fairer skin at an early stage on the other hand it is easily classifiable for an orinonal darker skin .I have used a dataset of approx 133 images out of 28 images have been actually used for training the classifiers. Some of them are:

Fig 3: dataset

Some samples of normal skin images are also given that is used for training the classifiers

Fig 4: Normal skin images

3.1 Algorithm:

- Prepare a training set of normal skin images
- Take input the original disease image.
- Convert it into binary image by taking care of the noise present in the image.
- Skeletonize the image.
- Extract the feature from the image
- Generate the SVM Classifiers and k-NN classifier the given image
- Compare the disease image against the training set
- Generate the result of the comparison

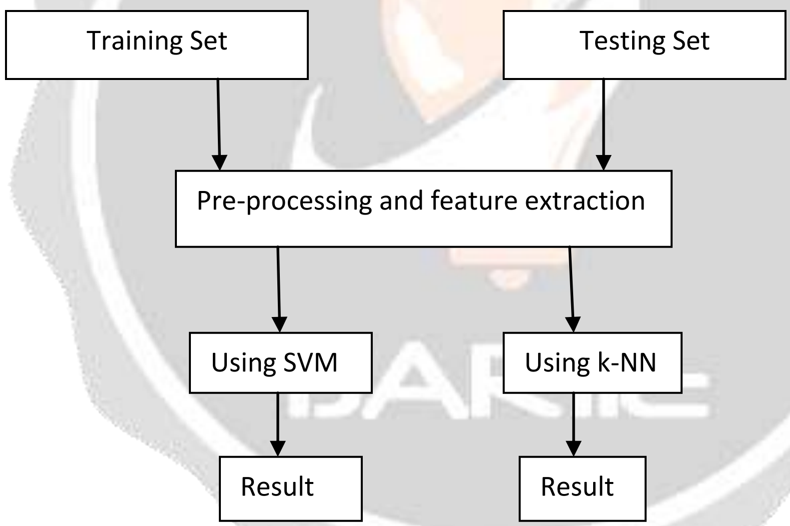

Fig 5: Block diagram of pre-processing of image

4. RESULT

We run both of the classifiers on different categories of data and compare their performances in terms of accuracy. For both of the classifiers, we can find Support Vector Machine classification accuracy rate is always the highest while k-Nearest Neighbour classification accuracy rate is always the lowest. One thing we want to mention here is the KNN learning curve does not seem that happy and the accuracy rate has not changed greatly .Some samples that I have used are given there. Training Set

Fig 6: Training set abnormal images

Fig 7: Training set normal images

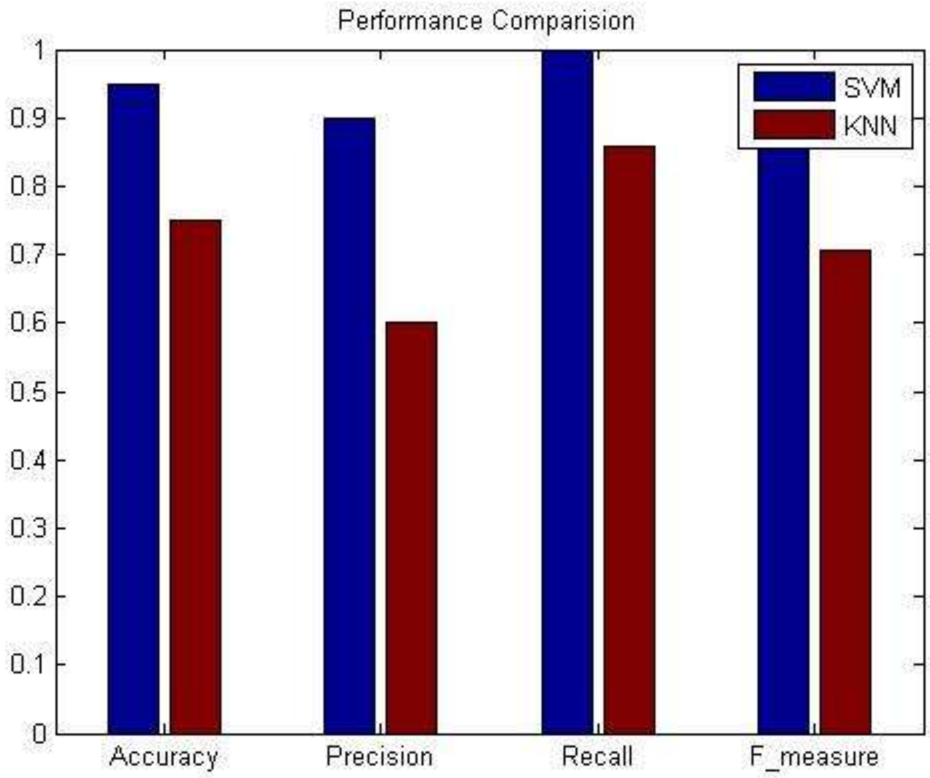

Fig 8: Performance rate of all feature types. a) KNN b) SVM

This implies the Support Vector Machine classification for leucoderma is more distinguishable than the other one. In the aspect of classifier performance on the same dataset, for our database, the k-NN algorithm performance is not as good as the SVM which guarantee us of meeting our first success metric Analysis: as the training set grows, the prediction quality increases, they are all happy graphs

4. CONCLUSIONS

As best of our knowledge there is no prior effort for Leucoderma detection and Analysis has been done. Hence, it is not possible to compare the results with others. On the other hand, SVM and k-NN is a novel techniques of data mining used for classification problems. I have compared the results of both the classification algorithms for the detection of leucoderma and found SVM is more accurate than k-NN that can easily be viewed by the graph. There are some images where it is quite difficult to recognize properly the existence of leucoderma for examples when the normal skin color is already fair. Therefore the average success rate we have obtained here is quite satisfactory.

5 REFERENCES

[1] Diane M.Thiboutot,"Dermatological Applications of High-Frequency Ultrasound ",Section of Dermatology, The Pennsylvania State University College of Medicine, Hershey, PA 17033

[2] M. F. Akay - Support vector machines combined with feature selection for breast cancer diagnosis. Expert Systems with Applications, 2009.

[3] S. Agatonovic-Kustrin and R. Beresford. Basic concepts of artificial neural network (ANN) modelling and its application in pharmaceutical research Journal of Pharmaceutical and Biomedical Analysis, 22:717–727, 2000.

[4] Michal Antkowiak. Recognition of skin diseases using artificial neural networks. In Proceeding of USCCS'05, pages 313–325, 2005.

[5] Chris M. Bishop. Neural networks and their applications. Review of Scientific Instruments, 65(6):1803–1832, 1994.

[6] Chris M. Bishop. Neural Networks for Pattern Recognition. Clarendon Press, Oxford, UK, 1995.

[7] A. Blum, H. Luedtke, U. Ellwanger, R. Schwabe, G. Rassner, and C. Garbe.Digital image analysis for diagnosis of cutaneous melanoma. Development of a highly effective computer algorithm based on analysis of 837 melanocytic lesions. British Journal of Dermatology, 151:1029–1038, 2004.

[8] E., Donald, "Introduction to Data Mining for Medical Informatics," Clin Lab Med, pp. 9-35, 2008.

[9] Irene M. Mullins et al., "Data mining and clinical data repositories: Insights from a 667,000 patient data set," Computers in Biology and

[10] J.C.Lobach, D. F.Goodwin, L. K.Hales, J. W.Hage, M. L. Edward Hammond, W. Parther, "Medical Data Mining: Knowledge Discovery in a Clinical Data Warehouse," 1997.

[11] Dan Dalan, "Clinical data mining and research in the allergy office," Current Opinion in Allergy & Clinical Immunology, vol. 10, no. 3, pp. 171-177, June 2010.

[12] Y. Mahajani, G. Aslandogan, "Evidence Combination in Medical Data Mining, "in Information Technology: Coding and Computing (ITCC), 2004.

[13]Cristianini, N. & Shawe-Taylor, J. (2000). An Introduction to Support Vector Machines and other kernel-based learning methods. Cambridge University Press, Cambridge, UK.

[14] Kanungo, T. Mount, D. M.Netanyahu, N. S. Piatko, C. D. Silverman, R.Wu, A.Y. (2002). "An efficient k-nearest neighbour algorithm: Analysis and implementation". IEEE Trans.Pattern Analysis and Machine Intelligence 24:881892. Doi: 10.1109/TPAMI.2002.1017616Retrieved2009-04-24.

[15] Kanungo, T.; Mount, D. M. Netanyahu, N. S.Piatko, C. D.Silverman, R.; Wu, A.Y. (2002). "An efficient k-nearest neighbour algorithm: Analysis and implementation". IEEE Trans.Pattern Analysis and Machine Intelligence 24:881892. doi: 10.1109/TPAMI.2002.1017616. Retrieved 2009-04-24.

[16] John C.Platt, Fast Training of Support Vector Machines using Sequential Minimal Optimization

[17] Shakhnarrovish, Darrell and Indyk, Nearest Neighbor Methods in learning and vision, MIT-press, 2005

[18] V. Vapnik - Statistical Learning Theory, Wiley, NewYork, 1998.

[19] John.C.Platt, N. Cristianini, and J. Shawe-Taylor. Large margin dags for multiclass classification. In Advances in Neural Information Processing Systems 12, pages 547–553, 2000.

[20] M.Vidyasagar. A Theory of Learning and Generalization.Springer-Verlag, New York, 1997.

[21] X. Zhou and B. Bhanu, “Feature fusion of face and gait for human recognition at a distance in video,” in Proc. IEEE Int. Conf. Pattern Recognition, vol. 4, pp. 529– 532, 2006.

[22] J.S.Suri, D.L.Wilson, Laxminarayan Seds. Handbook of biomedical image analysis, Vol. II. London: Kluwer Academic, 2005: 1–57.

[23] B.Boser, I .Guyon, V. Vapnik: A training algorithm for optimal margin classifiers. In: Fifth Annual Workshop on Computational Learning Theory, Pittsburgh, ACM (1992) 144–152

[24] T. Dietterich, G. Bakiri.: Solving multiclass problem via error-correcting output code. Journal of Artificial Intelligence Research, Vol. 2 (1995) 263–286

[25]J.Platt, N.Cristanini, Shawe-Taylor, J.: Large margin DAGs for multiclass classification. Advances in Neural Information Processing Systems 12. MIT Press (2000) 543–557

[26] Songyot Nakariyakul and David Casasent, SVMC: Single-Class Classification with Support Vector Machines.

[27]http://saravananthirumuruganathan.wordpress.com/2010/05/17/a-detailed-introduction- to-k-nearest-neighbor-knn-algorithm/

[28] http://www.articlediary.com/article/what-is-leucoderma-causes-symptoms-treatment- of leucoderma-249.html.

[29] Archana Shekokar, Ramakrishna Ghubde, Pedigree Analysis and Cytogenetic Study In Vitiligo.