SVM VS KNN FOR CLASSIFICATION OF HISTOPATHOLOGICAL IMAGES OF VARICOSE ULCER

Ajitha K

Department of Computer Science and Engineering, Dr.Sivanthi Aditanar College of Engineering, Tiruchendur, India

ABSTRACT

This paper proposes a clinical tissue classification scheme for varicose ulcer classification using medical image processing. Among various leg ulcers varicose ulcer accounts for over 90% of all cases. The proposed work is a system which could be used to identify and classify varicose ulcer on any stage. In this paper, classification is divided into two parts: feature extraction, KNN and SVM classification. In feature extraction step, texture features like homogeneity, energy, entropy, contrast, mean, dissimilarity and variance are extracted. These extracted features are classified using K-Nearest Neighbor classifier and Support Vector Machine to differentiate the different stages of ulcer. Our proposed approach gives efficient performance rates of average sensitivity (72.46%), specificity (85.17%), and accuracy (80.95%) for KNN classifier and efficient performance rates of average sensitivity (77.94%), specificity (86.78%), and accuracy (78.84%) for SVM classifier. The result was to find the efficient classifier for classification of varicose ulcer among SVM and KNN based on their accuracies.

KEYWORDS

Varicose ulcer, Feature Extraction, Classification, K-Nearest Neighbor classifier, Support Vector Machine classifier.

1. INTRODUCTION

With the continuous development of science and technology, digital medicine and digital images are increasingly developing into society. The technologies developed in recent years, such as image processing and virtual reality technology, are slowly being applied in the medical field.

A varicose ulcer is a clinical pathology of localized damage to the skin and underlying tissue caused by diabetics, shear, or friction. Clinicians usually evaluate each varicose ulcer by visual inspection of the damaged tissues, which is an imprecise manner of assessing the ulcer state. Current computer vision approaches do not offer a global solution to this particular problem. Histopathology refers to the microscopic examination of tissue in order to study the manifestations of disease. Specifically, in clinical medicine, histopathology refers to the examination of a biopsy or surgical specimen by a pathologist, after the specimen has been processed and histological sections have been placed onto glass slides. Varicose veins may be tortuous, twisted, or lengthened veins that are present in 10%–40% of people aged 30–70 years. This common disease has been neglected because most patients have mild complaints including cosmetic embarrassment, leg aching, pruritus, and skin rashes. Only relatively few patients experience a varicosity-related severe complication, such as a lower limb ulceration [1]. In a US community study, 5% of adults had skin changes in the lower extremities. It is estimated that 15 - 20% of Indian population is suffering from varicose veins. Varicose veins are more common in certain occupations like teachers, labourers, nurses, etc. Another study claims that 20- 25% women and 10-15% men have visible Varicose Veins [2]. One of the surveys on railroad workers showed that those in North India (6.8%) were less likely to develop Varicose Veins compared to those in the south (25.08%)[3].

Mingqiang Yang et al., KidiyoKpalma et al., Joseph Ronsin et al., [4] proposed that varicose ulcers are defined as open lesions between knee and the ankle joint that occur in the presence of varicose disease. They are the most common cause of leg ulcers, accounting for 60- 80% of them. The prevalence of VLUs is between 0.18% and 1%. Over the age of 6, the prevalence increases to 4%. On an average 33-60% of these ulcers persist for more than 6 weeks and are therefore referred to as chronic VLUs. These ulcers represent the most advanced form of chronic varicose disorders like varicose veins and lipodermatosclerosis.

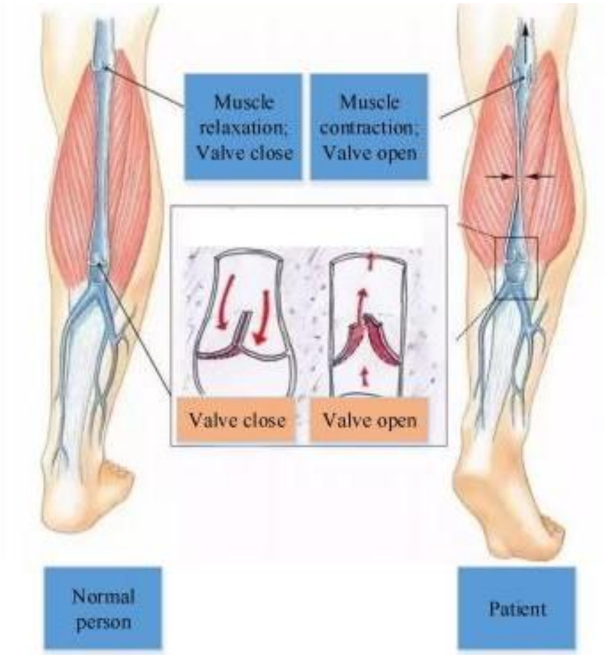


Figure 1: Comparison of varicose veins and normal veins

2. RELATED WORK

Zheng et al., Bradley et al., Patterson et al., Galushka et al., Winder et al., [5] proposed an efficient tissue classification for wound healing assessment. They presented a new tissue classification protocol using the RGB (Red, Green and Blue) histogram distributions of pixel values from wound color images. These three histogram distributions (extracted features) were used as three two- dimensional (2D) input signals for classification. This protocol has been carried out using the KNN classifier and results show that the proposed protocol provides an extremely competent practical method for the classification of wound tissues.

Green et al.,Pooler et al., McKinley et al., [7] proposed that varicose ulcers are common, intractable and often recurrent but care is often wound focused, potentially overlooking the significant impact the condition has on the patients' daily life. A systematic review was undertaken to explore the factors that impact on the quality of life of the patient with varicose ulcer. Twenty-three studies (11 qualitative / 12 quantitative) met the inclusion criteria and were the subject of full review. The qualitative studies were collapsed into four core themes: physical, psychological, social implications and the nurse-patient relationship. The quantitative studies were grouped according to the tool applied.The review demonstrated that varicose ulcers impact negatively across all areas of daily living. Pain, exudate, odour and the impact on mobility were daily challenges. The ability to engage with everyday functioning was restricted either due to the ulcer, the dressing or due to a self-imposed isolation in response to the impact of symptoms. Depression and low mood were common and yet, despite this, some studies reported that participants remained hopeful.

Nayak et al., addressed the composition of different types of tissue based on color and pigmentation inside the wound by image processing [8]. Previous works on tissue classification [9] have been carried out using SVM classifier with an accuracy rate of 87.61%. Misclassification is a common problem in any classification algorithm. Major works on tissue classification is based on the RGB histogram values [9].

Nejati, Hossein, et al., [10] In this they proposed shed light on fine-grained tissue classification to better realign the goal with clinically approved practices. They then presented approach to classify all 7 different tissue types, based on using a pre-trained DNN as a feature extractor for wound tissue classification. Then used DNN layers as image representation features and then perform feature reduction and classification using PCA and linear SVM, to reach patch-level labeling of the wound image. The proposed method not only outperforms previously proposed features, it is more robust in discrimination of similar looking tissue types and also against illumination condition changes.

3. PROPOSED SYSTEM

The main purpose of the proposed scheme helps to identify the ulcer in the initial stage and provides an efficient method to classify the ulcer and to start the treatment in appropriate time. The system includes the following stages, feature extraction and classification. The figure 1 shows system architecture, which shows the process of the proposed system. Texture features are extracted from the image. These extracted features are classified using K-Nearest Neighbor classifier and Support Vector Machine classifier to differentiate the different stages of the ulcer. The following section describes the modules involved in the proposed work.

4. METHODS

4.1. Feature Extraction

Feature Extraction is one of the most important tasks in image processing area which allows determining the most relevant features for pattern recognition. An appropriate subgroup of features is found when it allows the production of the similarity of the pattern within its class and dissimilarity among other different classes. Texture features are extracted from the ulcer of the segmented image.

4.1.1. Texture Features

Texture is about the spatial arrangement of color or intensities in an image or selected region of an image. Image textures can be assumed in natural scenes captured in an image. Image texture is used to help in segmentation of images [11].

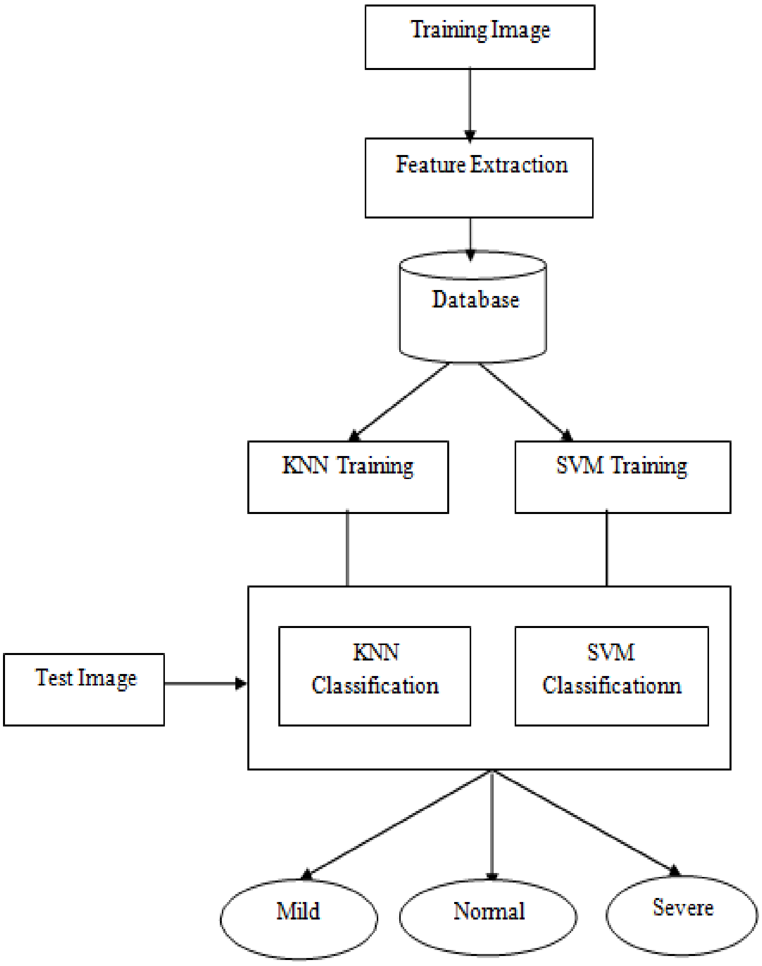


Figure 2: Block Diagram of the Proposed Work

- **Homogeneity**

Homogeneity can be defined as the measurement of degree of variance. It reflects the homogeneity of image textures and scaled the local changes of image texture. High values of homogeneity denote the absence of intra-regional changes and locally homogenous distribution in image textures.

$$\sum_{i,j-0}^{N-1} \frac{P_{i,j}}{1+(i-j)^2}$$

- **Energy**

Energy can be defined as the measure of the extent of pixel pair repetitions. It measures the uniformity of an image. When pixels are very similar, the energy value will be large.

$$\sum_{i,j-0}^{N-i} (P_{ij})^2$$

- **Entropy**

Entropy is the measure of randomness that is used to characterize the texture of the input image. Its value will be maximum when all the elements of the co-occurrence matrix are the same.

$$\sum_{j,j=0}^{N-1} -ln(P_{ij})P_{ij}$$

- **Contrast**

Contrast is to find the intensity between a pixel and its neighbor over the whole image. It is also used for enhancing the image quality. Contrast is the variation in color that makes an object distinct. Visually, contrast is identified by the difference in the color and brightness of the object and other objects within the same field of view.

$$\sum_{i}^{N} \sum_{j}^{N} (i - j)^2 P(i, j)$$

- **Mean**

Mean is measured by calculating the average of all pixel values within the image. Mean value gives the contribution of individual pixel intensity for the entire image.

$$\mu = \sum_{j=1}^{N} j\ (P_{ij}) = \sum_{i=1}^{N} i\ (P_{ij})$$

- **Dissimilarity**

Numerical measure of how different two data objects range from 0 (objects are alike) to ∞ (objects are different).

$$\sum_{i=1}^{N} \sum_{j=1}^{N} P_{ij} |i - j|$$

- **Variance**

Variance is normally used to find how each pixel varies from the neighbouring pixel (or centre pixel) and is used in classify into different regions.

$$V = V_i = \sum_{i=1}^{N} \sum_{j=1}^{N} (i - \mu_i)^2 (P_{ij})$$

4.2. Classification

4.2.1. KNN Classifier

K-nearest neighbor (kNN) classification [12] finds a group of k training tuples (k nearest neighbors) in the training set that are closest to the unknown tuple. To classify an unlabeled tuple, the distance of this unknown tuple to the labeled tuple is computed for identifying k- nearest neighbors and most common class labels of these nearest neighbors are then used to determine the class label of the unknown tuple. K-nearest neighbor algorithm (k-NN) is a

method of lazy learning. Classification of unknown tuples can be done using the closeness of unknown to the known according to some distance/similarity function. Euclidean distance is used as the distance metric. Euclidean distance between two tuples is estimated by:

$$dist(X_1, X_2) = \sqrt{\sum_{i=1}^{N} (x_{1i}, x_{2i})^2}$$

Once the nearest-neighbor list is obtained, the test tuple is classified based on the majority class of its nearest neighbors. If k = 1, then the unknown tuple is simply assigned the class of its nearest neighbor.

KNN Algorithm

Input: the set of training tuples and unlabeled test tuple.

Process: Compute the distance between unlabeled test tuple and each training tuple. Select the set of closest (k nearest neighbor) training tuples to the unlabeled tuple.

Output: label the test tuple with the majority class of its nearest neighbour.

4.2.2. SVM Classifier

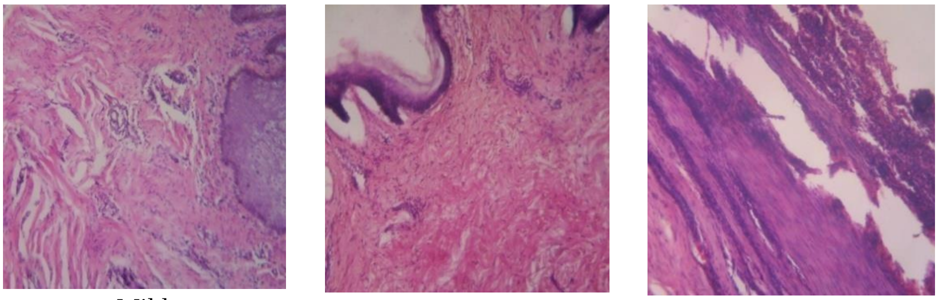
SVM is one of the most popular supervised machine learning methods used in neuroimaging, and it deals effectively with high dimensional data and it will provide a good classification results.The primary aim of a the SVM classifier is to identify the decision surface that would distinguish between classes and based on that surface other new and unseen data instances are assigned into the groups.In the training phase, the classifier computes the optimal decision surface expressed in the form f(x) = w•x + b4 only by a subset of the original training set D=< $\boldsymbol{xi}$, $\boldsymbol{yi}$ > called the support vectors. Support vectors are nothing but the data points that lie closest to the optimal separating hyperplane and hence are the most difficult patterns to classify.

The optimal hyperplane is determined by maximizing the margin of separation between the two classes. The part of this quadratic problem is to ensure that no data points can lie in the margin.

In the testing phase, the classifier is required to predict the label $yi$of new, previously unseen data, by evaluating y = sgn(w•x + b).In some cases data are not linearly separable,so kernels are introduced to the machine.Kernels are functions that allow a mapping of the original, non-linearly separable data into a new feature space where the data are linearly separable. Polynomial, Gaussian and radial basis function (RBF) are some of the most commonly used kernels.

5. EXPERIMENTAL RESULTS

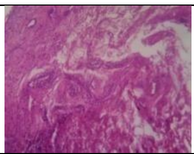
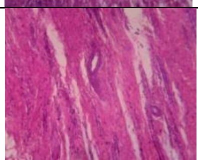
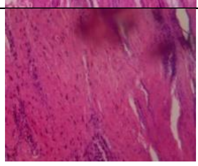
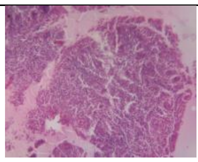
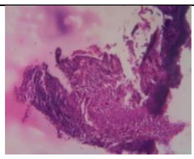
The various varicose ulcer images were collected from various hospitals and studied under clinical guidance. Totally 33 histopathological images of varicose ulcer are broadly classified as normal, mild and severe were collected and trained. Texture Feature descriptors were extracted from the image. KNN and SVM classifier was trained. The images were trained to the classifier by specifying the type of stages such as type 1 for mild ,type 2 for normal and type 3 for severe. For the test samples the feature extraction step was executed. Finally the difference in stages of ulcer was classified using KNN classifier and SVM Classifier. Figure 3 shows the sample varicose ulcer images in each stage.



Mild Normal Severe

Figure 3: Sample Varicose Ulcer Images

Table 1: Intermediate results of the proposed work.

Input image	Clinicians diagnosed wound stage	Propose d work
	Mild	Mild
	Mild	Normal
	Normal	Normal
	Severe	Severe
	Severe	Severe

5.1. Performance Assessment Metrics

Percentage of correctly identified ulcer stage that is, True positive is used to find the Sensitivity of the proposed method.

$$\text{Sensitivity} = \frac{\text{No of cases matched with clinician s report}}{\text{No of cases matched with clinicians report} + \text{No of cases not matched with clinicians report}} \times 100$$

The specificity is evaluated based on false positive predictions. False positive rate is the percentage of ulcer stages identified by our system but not reported by clinicians. In rare cases, the specificity would be useful to detect the correct stage misinterpreted by clinicians.

$$\text{Specificity} = 1 - \frac{\text{No of cases identified by the system}}{\text{No of cases reported by the clinicians}} \times 100$$

Accuracy is the measure which gives the predictive ability of the proposed system in terms of True positive, True negative, False positive, and False negative values.

$$\text{Accuracy} = \frac{\text{True positive} + \text{True negative}}{\text{True positive} + \text{True negative} + \text{False positive} + \text{False negative}}$$

These three different measures shown in are used to quantify the proposed system. A total of 33 images were involved to train the system. The test sets are evolved based on the random sampling method and 30 images separately were used for testing. The NetBeans IDE 8.2 was used to implement the methodology. The training and testing were carried out in the 8.1 PC configured by Intel Core I3 CPU 2.3 GHz, 2 GB RAM and 250 GB storage. The confusion matrix for each test set is derived to calculate the performance. Table 2 and 3 shows the confusion matrix for the test set executed by KNN and SVM classifier.

Table 2: Confusion matrix for the test set executed by KNN classifier.

Test set	True Positive	False Positive	True Negative	False Negative
T1	15	6	35	7
T2	14	7	38	4
T3	15	6	36	6

Table 3: Confusion matrix for the test set executed by KNN classifier.

Test set	True Positive	True Negative	False Positive	False Negative
T1	11	10	39	3
T2	11	10	42	0
T3	21	0	25	17

Table 4. Statistical performance analysis of KNN and SVM Classifier.

Test set	KNN Classifier		SVM Classifier	
	Sensitivity	Specificity	Sensitivity	Specificity
T1	0.681818	0.853659	0.785714	0.795918
T2	0.777778	0.844444	1	0.807692
T3	0.714286	0.857143	0.552632	1

Table 5. Accuracy analysis of KNN and SVM Classifier.

Test set	KNN Classifier	SVM Classifier
	Accuracy	Accuracy
T1	0.793651	0.793651
T2	0.825397	0.84127
T3	0.809524	0.730159

The above stated performance measures are calculated based on the results of the testing phase. Figure 4 shows the pictorial sensitivity specificity analysis of KNN and SVM classifier. The accuracy analysis results are shown in Table 5. While experimenting the results the proposed methods KNN Classifier give the sensitivity (72.46%), specificity (85.17%), accuracy (80.95%) and SVM classifier give the sensitivity (77.94%), specificity (86.78%), and accuracy (78.84%).

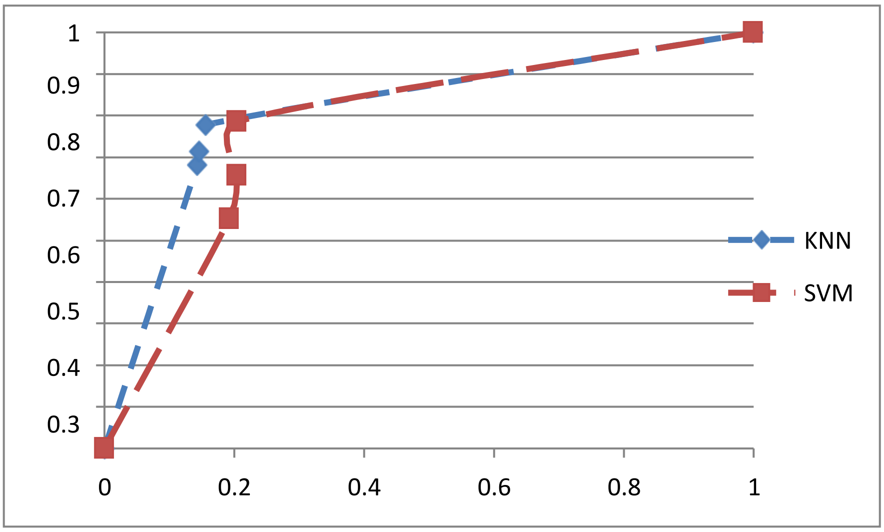


Figure 4. The pictorial sensitivity specificity analysis of KNN and SVM classifier.

CONCLUSION

Computer aided image analysis for varicose ulcer tissue classification is much needed nowadays for early diagnosis and treatment. The proposed model possesses texture feature extraction which would be very useful in identifying the stages of the wound. This method also includes two classifier so the accuracy from both the classifier would help us to find a better classifier for future works. In future the same methodology may be carried with more number of datasets and many more processing steps so that the result would be more accurate for the classifiers to identify the stages of the ulcer.

REFERENCES

[1] Golledge J, Quigley FG. Pathogenesis of varicose veins. Eur J Vasc Endovasc Surg 2003;25:319–324.

[2] Callam MJ. Epidemiology of varicose veins. Br J Sur 1994; 81: 167-73.

[3] Malhotra SL. An epidemiological study of varicose veins in Indian railroad workers from south and north of India, with reference to the causation and prevention of varicose veins. Int J Epid 1972; 1: 177- 83.

[4] Bhuptani, N. and Talati, B., 2014. An Efficient Image Retrieval Technique using Shape Context Feature. International Journal of Computer Applications, 975, p.8887.

[5] Zheng, H., Bradley, L., Patterson, D., Galushka, M. and Winder, J., 2004, September. New protocol for leg ulcer tissue classification from colour images. In The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (Vol. 1, pp. 1389-1392). IEEE.

[6] Galushka M, Zheng H, Patterson D, et al. Case-based tissue classification for monitoring leg ulcer healing. Proceedings of the 18th IEEE Symposium on Computerbased Medical Systems; 2005.

[7] Green, J., Jester, R., McKinley, R. and Pooler, A., 2014. The impact of chronic venous leg ulcers: a systematic review. Journal of wound care, 23(12), pp.601-612.

[8] Nayak R, Kumar P, Galigekere RR. Towards a comprehensive assessment of wound-composition using color-image processing. Proceedings of the IEEE International Conference on Image Processing (ICIP ‘09); 2009 Nov. p. 4185–4188.

[9] Mukherjee R, Manohar DD, Das DK, et al. Automated tissue classification framework for reproducible chronic wound assessment. 2014 Jul 8; doi:10.1155/2014/851582. PMCID:PMC4121018

[10] Nejati, Hossein, et al. "Fine-grained wound tissue analysis using deep neural network." 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018.

[11] Marsilin JR,. Wiselin Jiji G. An efficient CBIR approach for diagnosing the stages of breast cancer using KNN classifier. Bonfring International Journal of Advances in Image Processing. 2012;2(1):1

[12] P-N. Tan, M. Steinbach, V. Kumar, ―Introduction to data mining‖, Pearson Addison-Wesley, 2006

[13] Bhavani, R.R. and Wiselin Jiji, G., 2018. Image registration for varicose ulcer classification using KNN classifier. International Journal of Computers and Applications, 40(2), pp.88-97.

[14] Bremner D, Demaine E, Erickson J, Iacono J, Langerman S, Morin P, Toussaint G, Outputsensitive algorithms for computing nearest neighbor decision boundaries, Discrete and Computational Geometry, 2005, pp. 593–604.

[15] Cover, T., Hart, P. , Nearest-neighbor pattern classification, Information Theory, IEEE Transactions on, Jan. 1967, pp. 21-27