VISVESVARAYA TECHNOLOGICAL UNIVERSITY

JNANA SANGAMA, BELAGAVI - 590014

A Synopsis Report

On

" Support Vector Machine for Assistant Clinical Diagnosis of Cardiac Disease "

Submitted in the partial fulfillment of the requirement for the award of

Bachelor of Engineering

In

Computer Science and Engineering

Submitted By

SURYA GANGARAJ K **(1DT15CS115)**

Under the guidance of

Dr.R.Saravana Kumar

Associate Professor

Dep. of CSE, DSATM

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Accredited 3 years by NBA, New Delhi (Validity : 26-07-2018 to 30-06-2021)

DAYANANDA SAGAR ACADEMANY OF TECHNOLOGY & MANAGEMENT

Udayapura, Kanakapura Main Road, Opp. Art of Living, Bangalore-82

2018 -2019

INTRODUCTION

Support-vector machines (SVM)[3] are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. SVM is powerful for the characterized by small sample, nonlinearity, high dimension and local minima. SVM implements well trade off between the quality of the approximation of the given data and the complexity of the approximating function by Structural Risk Minimization (SRM) principle. Currently SVM[3] is an active field in artificial intelligent technology, and has been applied to pattern recognition, function estimation, signal processing, control and other fields .

The paper is organized as follows. SVM nonlinear classification algorithm is reviewed in Section II and parallel decision model based on SVM[3] is introduced in Section III. In Section IV, SVM is applied to Clinical Diagnosis of Cardiac Disease. Conclusions are given in Section V.

Literature review

PROS	CONS
SVM have are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis	Isn't suited to larger datasets as the training time with SVMs can be high
SVM have higher Accuracy	Less effective on noisier datasets with overlapping classes
It is useful for both Linearly Separable(hard margin) and Non-linearly Separable(soft margin) data. The only thing to do is to come up with the optimal penalty variable C (the one that multiplies slack variables)	In Natural Language Processing, structured representations of text yield better performances. Sadly, SVMs[3] can not accommodate such structures(word embeddings) and are used through Bag-of-Words representation which loses sequentiality information and leads to worse performance.
Conformity with Semi-Supervised Learning: It may be used in a dataset where some of the data are labeled and some are not. You only add an additional condition to the minimization problem and it is called Transductive SVM.	Edit from Shuyang Sheng: SVM in its vanilla form cannot return a probabilistic confidence value like logistic regression does, in some sense it's not 'explanatory' enough. Confidence of prediction can be quite important in many applications.

Feasibility Study

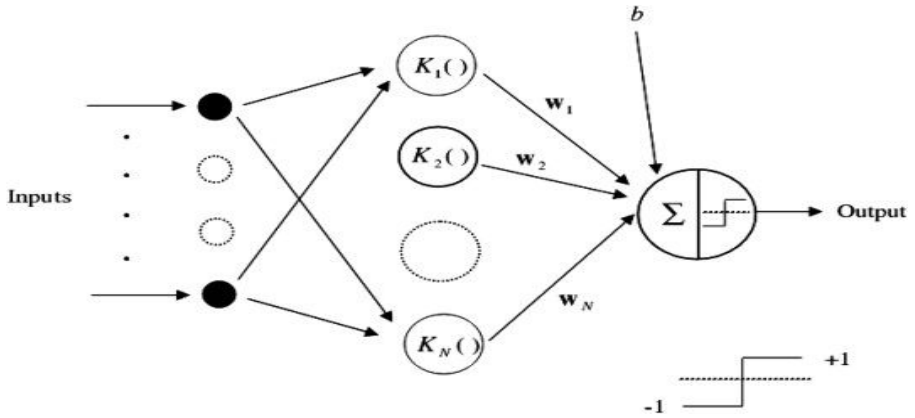
SVM have always given better results than other machine learning models, The machine learning for medical use should have most accurate results, so two cases are studied. SVm for the detection of Microcalcifications and SVM for the detection of Tuberculosis infection.
In the case of Microcalcifications, SVM classifier was trained through supervised learning to test at every location in a mammogram whether an MC is present or not. The formulation of SVM learning is based on the principle of structural risk minimization. The decision function of the trained SVM classifier is determined in terms of support vectors that were identified from the examples during training. The result is that the SVM classifier achieves low generalization error when applied to classify samples that were not included in training
In the case of Tuberculosis detection, SVM have performed well and has advantages which indicate it may be the best approach for the final classification of image objects.These hypothesis modeling the data is formulated in terms of the most informative patterns (SVM)

Methodology

SVM is one of the best model in the market which gives better output or result on lesser training time. The electrocardiogram is today used worldwide as a relatively simple way of diagnosing heart conditions. An ECG is a recording of the small electric waves being generated during heart activity. An ECG intervals and waves are shown in Fig. 2. This diagram illustrates ECG waves and intervals as well as standard time and voltage measures on the ECG paper.

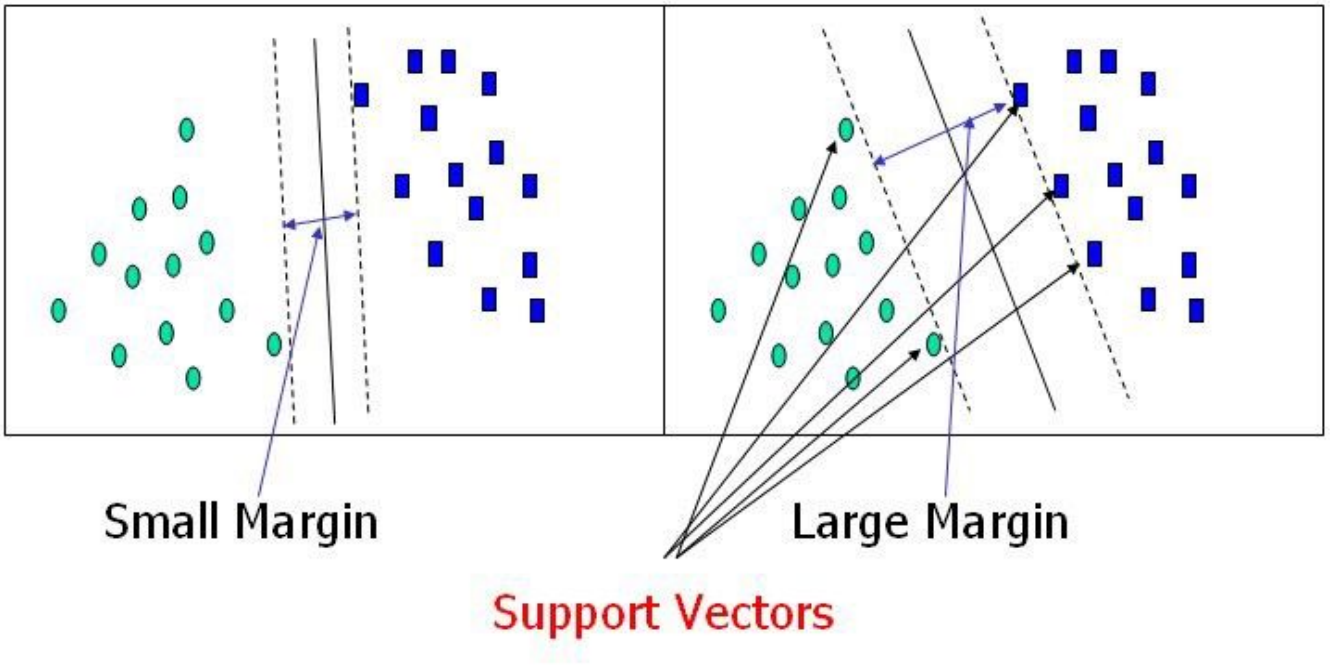
The P wave represents atrial activation; the PR interval is the time from onset of atrial activation to onset of ventricular activation. The QRS complex represents ventricular activation; the QRS duration is the duration of ventricular activation. The ST-T wave represents ventricular repolarization. The QT interval is the duration of ventricular activation and recovery. The U wave probably represents "after depolarizations" in the ventricles.

ECG examination is one of means in clinical diagnosis, which is important to some diseases especially cardiac disease. Using computer to identify and classify ECG, first divide ECG exactly, and extract every wave, finally analyse data. This "method" is recommended when reading all 12-lead ECG's. Like the physical examination, it is desirable to follow a standardized sequence of steps in order to avoid missing subtle abnormalities in the ECG tracing, some of which may have clinical importance .



The data are from LDS Hospital in Salt Lake City provided by Frank G.Yanowitz of University of Utah School of ECG Department. Most of the 12- and 6- lead ECGs were recorded at LDS Hospital in Salt Lake City, Utah. Marquette Electronics has also given permission to use ECG rhythms and diagrams from their educational posters. Each of the ECGs has an interpretation and many have additional explanations that help explain the diagnosis .

There are 8 leads after the transform of ordinary 12- lead ECG. Every one notes the signal of ECG in 10 seconds. In the original data, every lead include 5000 samples, so every swatch has (8 × 5000) data. For the difference of individual and heartbeat frequency, the extracted full heartbeat samples are difference from one another. For the effective pattern classification.



796 samples act as a training set, and 400 samples act as a testing set to evaluate learning result from LDS Hospital database. All experiments are implemented on Pentium 350 MHz with 512 MB RAM. Matlab 6.5 is employ to solve the quadratic programming.

Bibliography/References

[1]A Support Vector Machine Approach for Detection of Microcalcifications
Issam El-Naqa, Student Member, IEEE, Yongyi Yang*, Member, IEEE, Miles N. Wernick, Senior Member, IEEE, Nikolas P. Galatsanos, Senior Member, IEEE, and Robert M. Nishikawa
[2] The applications of machine learning for the detection of tuberculosis infection by Veropoulos, N.Cristianini .U.K
[3] SVM model Google scholar
[4] Geeks4Geeks.com