VISVESVARAYA TECHNOLOGICAL UNIVERSITY
BELGAUM – 590 014

A Project Seminar Report on

"Machine Learning for the Diagnosis of Disease"

A seminar report submitted in partial fulfillment of the requirements for the reward of the degree of ***Bachelor of Engineering in Computer Science and Engineering*** *of Visweswaraya Technological University, Belgaum.*

Submitted by:
SURYA GANGARAJ K (1DT15CS115)

Under the guidance of:
Dr.R.Saravana Kumar
(Associate Professor,Dept. of CSE)

Department of Computer Science and Engineering

DAYANANDA SAGAR ACADEMY OF TECHNOLOGY AND MANAGEMENT

Udayapura, Kanakapura Road, Bangalore-560082

2018-2019

DAYANANDA SAGAR ACADEMY OF TECHNOLOGY AND MANAGEMENT,

Udayapura, Kanakpura Road, Bangalore

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

CERTIFICATE

This is to certify that the project seminar work entitled **“An Application of Machine Learning in the Diagnosis of Ischaemic Heart Disease”** is carried out by **SURYA GANGARAJ K (1DT15CS115)** in partial fulfilment for the award of the degree of **Bachelor of Engineering** in **Computer Science and Engineering** of the **Visvesvaraya Technological University, Belgaum** during the year 2018-2019.The project seminar report has been approved as it satisfies the academic requirements with respect to the work prescribed for the award of Bachelor of Engineering Degree.

Signature of the Guide	Signature of the Seminar Coordinator	Signature of the HOD
(Dr.R.Saravana Kumar)	**(Prof. BHASKER RAO)**	**(Dr. C. NANDINI)**

ACKNOWLEDGEMENT

I express my appreciation for **Dr. B. R. Lakshmikantha**, **Principal**, DSATM, Bengaluru, in providing necessary facilities and for encouragement to carry out the Seminar.

I express my humble gratitude and thanks to **Dr. C. Nandini**, **Professor and HOD**, Computer Science and Engineering, DSATM for guiding us and having facilitated us to complete the seminar successfully.

I am very thankful to my guide **Dr. R. Saravana Kumar,** Associate professor, Department of Computer Science and Engineering**,** DSATM for their constant encouragement, guidance and support throughout the course of the project & Seminar without whose efforts this would not have been possible.

I thank the Seminar Coordinator **Prof. Bhasker Rao**, Associate Professor, Department of Computer Science & Engineering, DSATM, Bengaluru for their invaluable critical reviews, guidance and support.

I finally thank my project team and friends who have been encouraging constantly and inspiring throughout.

SURYA GANGARAJ K

USN: 1DT15CS115

ABSTRACT

In the past few decades, machine learning tools have been successfully used in several medical diagnostic problems. While they often significantly outperform expert physicians (in terms of diagnostic accuracy, sensitivity and specificity), they are mostly not being used in practice. One reason for this is that it is difficult to obtain an unbiased estimation of the diagnosis's reliability. We discuss how the reliability of diagnoses is assessed in medical decision-making and propose a general framework for reliability estimation in machine learning, based on transductive inference. We compare our approach with the usual machine-learning probabilistic approach, as well as with classical step-wise diagnostic process, where the reliability of a diagnosis is presented as its post-test probability. The proposed transductive approach is evaluated in a practical problem of the clinical diagnosis of Ischemic heart disease. Significant improvements over existing techniques are achieved.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Ischemic heart disease (IHD) is the most important cause of mortality in developed as well as in developing countries. Therefore improvements as well as the rationalization of diagnostic procedures and treatment of IHD are necessary. The usual procedure in IHD diagnosis consists of four diagnostic levels containing evaluation of signs and symptoms of the disease and ECG at rest, sequential ECG testing during a controlled exercise, myocardial scintigraphy and coronary angiography as a final test. Because the suggestibility is possible, the results of each step are interpreted individually and only the results of the highest step are valid. The amount of data available for each patient in all diagnostic levels is too large to be efficiently and objectively evaluated by physicians. The goal of a rational diagnostic algorithm is to establish the conclusive diagnosis of IHD and to plan the most appropriate management of the disease using only the necessary diagnostic steps.

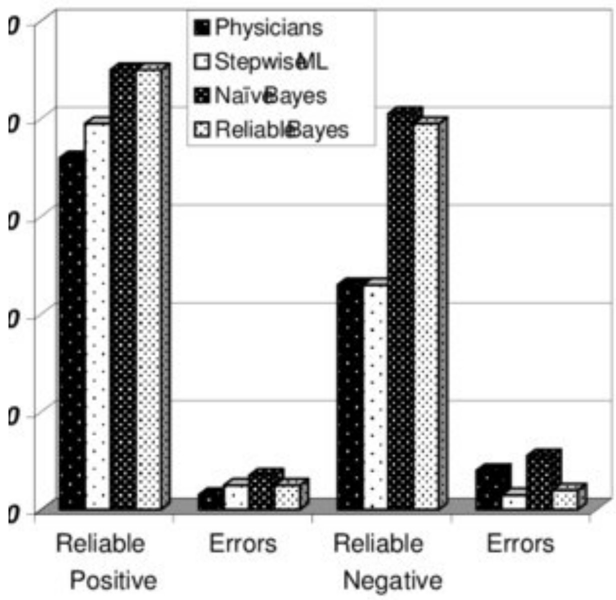
This can be achieved by taking into account and evaluating all the information collected by different diagnostic methods according to their importance and diagnostic value.

The performance of different diagnostic methods is usually described as classification accuracy (in Machine Learning) or as sensitivity and specificity (in medicine):

accuracy = (true positive+true negative test results) / all patients

sensitivity = true positive test results / all patients with disease

specificity = true negative test results / all patients without disease



CHAPTER 2

LITERATURE SURVEY

The function of the heart is pumping blood to all the organs of the body. For this task an uninterrupted supply of oxygen to the heart muscle is needed. This is achieved by sufficient blood flow trough the coronary arteries to the heart muscle - myocardium. In case of diminished blood flow through coronary arteries due to stenosis or occlusion, IHD develops. The consequence of IHD is impaired function of the heart and lastly necrosis of the myocardium myocardial infarction.

During the exercise the blood flow through the body has to be increased. Therefore the delivery of oxygen to the heart muscle has to increase several times by increasing blood flow trough the coronary arteries. In a (low grade) IHD the blood flow is sufficient at rest or during a moderate exercise, as perfusion of the myocardium is adequate, but insufficient during a severe exercise. Therefore, signs and symptoms of the disease develop only then.

There are four levels of diagnostics of IHD. First signs and symptoms of the disease arc evaluated clinically and ECG is performed at rest. This is followed by the sequential ECG testing during controlled exercises. If this test is not conclusive, or if additional information regarding the perfusion of the myocardium is needed, myocardial scintigraphy is performed. Radioactive material which accumulates in the heart muscle proportionally to its perfusion is injected into the patient and the images (scintigrams) showing perfusion of the heart muscle during exercise and rest are taken. By comparing both sets of images, the presence, localization and distribution of the ishaemic tissue arc determined. If an invasive therapy of the disease is contemplated, i.e. coronary artery bypass surgery, the diagnosis has to be concluded by imaging of the coronary vessels (injecting the contrast material into the coronary vessels and imaging their anatomy by x-ray coronary angiography).

In our study we used a dataset of 327 patients with performed clinical and laboratory examinations, exercise ECG, myocardial scintigraphy and coronary angiography. In 229 cases the disease was angiographically confirmed and in 98 cases it was excluded. The patients were selected from the population of approximately 4000 patients who were examined at the Nuclear Medicine Department in years 1991-1994. For the sake of our study we selected only the patients with complete diagnostic procedures (all four levels). Our experiments were conducted on four problems, depending on the amount of clinical and laboratory data (attributes) available for learning (between **30** and 77 attributes)

CHAPTER 3

PROPOSED SYSTEM OVERVIEW AND CONSTRUCTION

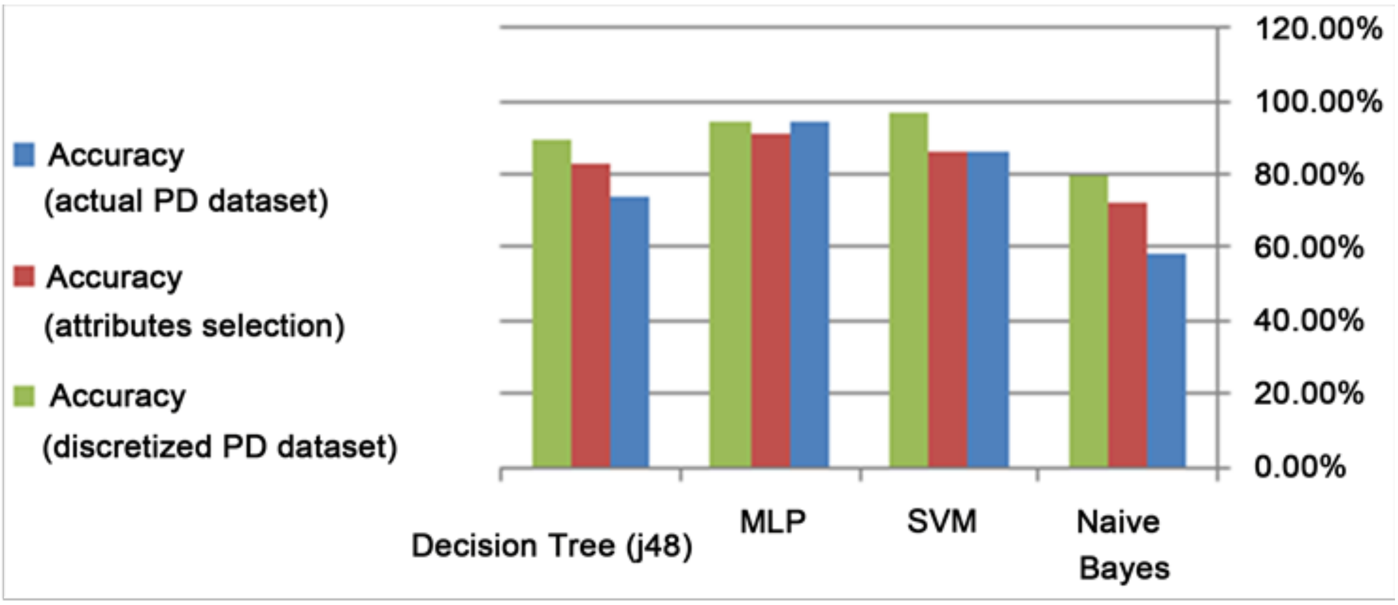
The function of the heart is pumping blood to all the organs of the body. For this task an uninterrupted supply of oxygen to the heart muscle is needed. This is achieved by sufficient blood flow trough the coronary arteries to the heart muscle - myocardium. In case of diminished blood flow through coronary arteries due to stenosis or occlusion, IHD develops. The consequence of IHD is impaired function of the heart and lastly necrosis of the myocardium myocardial infarction.

During the exercise the blood flow through the body has to be increased. Therefore the delivery of oxygen to the heart muscle has to increase several times by increasing blood flow trough the coronary arteries. In a (low grade) IHD the blood flow is sufficient at rest or during a moderate exercise, as perfusion of the myocardium is adequate, but insufficient during a severe exercise. Therefore, signs and symptoms of the disease develop only then.

There are four levels of diagnostics of IHD. First signs and symptoms of the disease arc evaluated clinically and ECG is performed at rest. This is followed by the sequential ECG testing during controlled exercises. If this test is not conclusive, or if additional information regarding the perfusion of the myocardium is needed, myocardial scintigraphy is performed. Radioactive material which accumulates in the heart muscle proportionally to its perfusion is injected into the patient and the images (scintigrams) showing perfusion of the heart muscle during exercise and rest are taken. By comparing both sets of images, the presence, localization and distribution of the ishaemic tissue arc determined. If an invasive therapy of the disease is contemplated, i.e. coronary artery bypass surgery, the diagnosis has to be concluded by imaging of the coronary vessels (injecting the contrast material into the coronary vessels and imaging their anatomy by x-ray coronary angiography).

In our study we used a dataset of 327 patients with performed clinical and laboratory examinations, exercise ECG, myocardial scintigraphy and coronary angiography. In 229 cases the disease was angiographically confirmed and in 98 cases it was excluded. The patients were selected from the population of approximately 4000 patients who were examined at the Nuclear Medicine Department in years 1991-1994. For the sake of our study we selected only the patients with complete diagnostic procedures (all four levels). Our experiments were conducted on four problems, depending on the amount of clinical and laboratory data (attributes) available for learning (between **30** and 77 attributes).

The **naive Bayesian classifier[2]** uses the naive Bayes formula to calculate the probability of each class given the values of all the attributes and assuming the conditional independence of the attributes. The attributes are usually defined by a human (often in medicine), and are therefore relatively independent, as humans tend to think linearly. This is the reason why the naive Bayesian formula often performs well on real-world problems.



CHAPTER 4

IMPLEMENTATION

In our experiments, the misclassification costs varied between 1 : 20 in favor of the “negative” class (no IHD present; higher specificity) and 20 : 1 in favor of the “positive” class (IHD present; higher sensitivity). The results of our experiments with some of the utilized algorithms are shown in Figures 1-**4** Each algorithm's behavior is shown in two figures. The first one depicts classification accuracy, information score, sensitivity and specificity. The vertical line marks the *uniform* cost (1 : 1) situation (

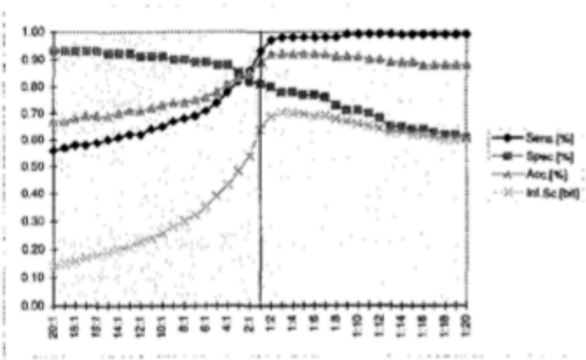


Figure 1. Naive Bayesian classifier

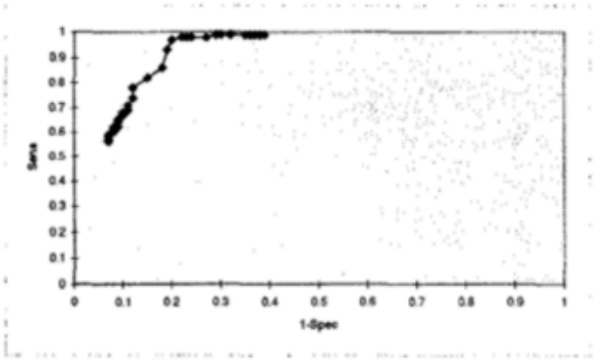


Figure 2. Naive Bayesian classifier - ROC curve

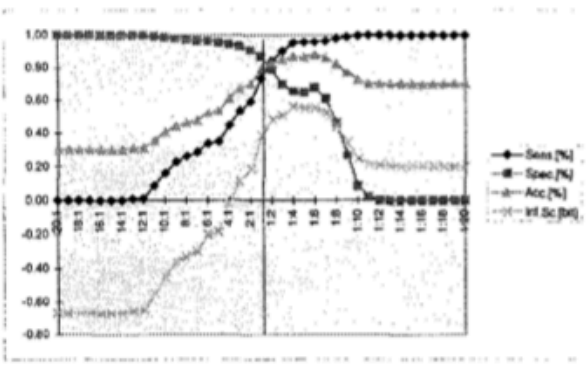


Figure 3. Neural net

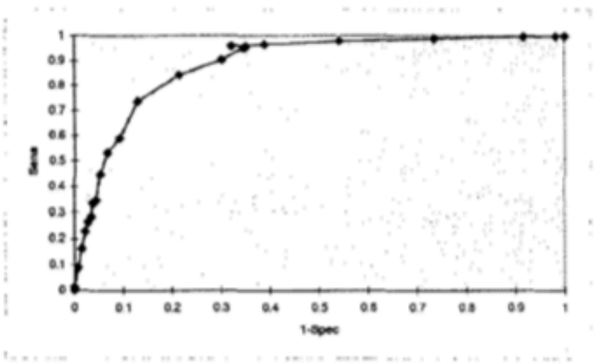


Figure 4. Neural net - ROC curve

behavior of the unmodified algorithm).

The second figure shows the ROC (relative operating characteristic) curve, that IS, a ratio between sensitivity and specificity. By changing the misclassification costs, one actually traverses along this curve. The results shown are averages of the ten-fold cross validation

CHAPTER 5

RESULTS

the Machine Learning algorithms are tuned to maximize classification accuracy. In our case, the sensitivity and specificity were much more important, so we generalized the algorithms to take in account the variable miss classification costs. The costs can be tuned in order to bias the algorithms towards higher sensitivity or specificity. We conducted many experiments with four learning algorithms and different variations of our dataset (327patients with completed diagnostic procedures). Our results show that improvements using Machine Learning techniques are reasonable and might find good use in practice.

	Naive Bayes			Neural net		
	Acc.	Sensit.	Specif.	Acc.	Sensit.	Specif.
1.	79.1	89.2	54.5	79.2	85.5	63.8
2.	79.7	89.3	57.1	81.6	88.5	65.0
3.	88.5	91.7	80.1	89.7	93.8	79.5
4.	87.3	90.2	80.1	88.4	92.6	79.1
	Assistant-I			Assistant-R		
	Acc.	Sensit.	Specif.	Acc.	Sensit.	Specif.
1.	71.2	73.4	59.3	73.2	76.1	61.9
2.	70.5	73.2	59.3	73.1	76.8	61.0
3.	89.0	89.1	88.1	86.6	89.6	79.7
4.	87.2	88.9	83.2	84.0	87.4	73.5

REFERENCES AND BIBLIOGRAPHY

1. IEEE Paper - https://ieeexplore.ieee.org/document/596411
2. Naive Bayes classifier - en.wikipedia.org/wiki/Naive_Bayes_classifier
3. NeuralNetworks - neuralnetworksanddeeplearning.com/chap2.html
4. Decision Tree - https://en.wikipedia.org/wiki/Decision_tree

1. Book - Machine Learning in Medicine - a Complete Overview
2. Book - Machine Intelligence for Healthcare