High Pitta Classification Using Finger Photoplethysmograph based features: A Feasibility Study

A Thesis submitted in partial fulfilment of the requirement for the award of degree of

Master of Engineering
in
Electronic Instrumentation and Control

Submitted By:
Bharti Chauhan
(801051005)

Under the supervision of:
Dr. Mandeep Singh
Associate Professor
EIED

July 2012

Electrical and Instrumentation Engineering Department
Thapar University

(Established under the section 3 of UGC act, 1956)
Patiala, 147004

Declaration

*I hereby declare that the report entitled "**High Pitta Classification Using Finger Photoplethysmograph based features: A feasibility study,**" is an authentic record of my own work carried out as a requirement for the award of degree of M.E. (Electronic Instrumentation & Control) at Thapar University, Patiala, under the guidance of **Dr. Mandeep Singh,** Associate Professor, Department of Electrical and Instrumentation Engineering, Thapar University during January to July 2012.*

Bharti Chauhan

Roll No. 801051005

Date: 12th July, 2012

It is certified that the above statement made by the student is correct to the best of my knowledge and belief.

12/7/12.

Dr. Mandeep Singh

Associate Professor, EIED

Thapar University, Patiala

Countersigned by:

S. Ghosh

Dr. Smarajit Ghosh

Professor & Head, EIED

Thapar University, Patiala

Dr. S.K. Mohapatra

Dean of Academic Affairs

Thapar University, Patiala

Acknowledgements

The real spirit of achieving a goal is through the way of excellence and austere discipline. I would have never succeeded in completing my task without the cooperation, encouragement and help provided to me by various personalities.

With deep sense of gratitude I express my sincere thanks to my esteemed and worthy supervisor, ***Dr. Mandeep Singh, Associate Professor, Department of Electrical & Instrumentation Engineering,*** *Thapar University, Patiala, for his valuable guidance in carrying out this work under his effective supervision, encouragement, enlightenment and cooperation.*

I shall be failing in my duties if I do not express my deep sense of gratitude towards ***Dr. Smarajit Ghosh, Professor & Head of the department of Electrical & Instrumentation Engineering,*** *Thapar University, Patiala who has been a constant source of inspiration for me throughout this work.*

I am also thankful to all the staff members of the Department for their full cooperation and help.

My greatest thanks are to all who wished me success especially my parents, family and friends who encouraged and supported at every point of time. Above all I render my gratitude to the ALMIGHTY who bestowed self-confidence, ability and strength in me to complete this work.

Place: Thapar University, Patiala

Date: 12th July 2012

Bharti Chauhan

Roll No. 801051005

Abstract

Ayurveda is the ancient Indian science of healing. Its primary diagnostic technique is through Nadi Parikshan i.e. pulse diagnosis. In an endeavour to revive this art, modern Instruments in the form of finger Photoplethysmography is employed by some researchers in their recent work. In a parallel study in this area, several features are extracted from the different fingers in a group of three subjects under consideration. As per Ayurveda the root cause of any disease is on account of imbalance of one or two of the three doshas namely Vata-Pitta-Kapha. It is also documented in at several places that Pitta naturally increases during the middle of life, the day, and the middle of digestion therefore the significant rise in the Pitta dosha is mainly in daytime i.e. in afternoon and after lunch. In light of this, several finger Photoplethysmographic waves were acquired from the subjects two hour before the lunch and within half an hour after the lunch. The features were analysed for the significant change on account of increased Pitta. This thesis work explores the statistically significant change in all the features so extracted and reports the features with a high feasibility to form the input vector for a high Pitta classifier.

Organization of Thesis

The complete project thesis is divided into seven chapters as follows.

The first chapter introduces Classification, classification rules, Steps of Classification and finally explains the various classification techniques.

The second chapter discusses the basics of the concepts used such as Ayurveda, Pitta dosha, Finger Photoplethysmograph and t-test.

The third chapter tells about the work that has been already carried out in this field.

The fourth Chapter formulates the problem.

The fifth chapter shows the result obtained in tabular form and discussion over the result.

In the seventh chapter thesis concluded with future scopes.

CONTENTS

List of Figures

List of Tables

Keywords

Classification

Ayurveda

Pitta

Finger Photoplethysmography

Features of Finger Photoplethysmography

t-test

Chapter-1

INTRODUCTION

1.1 Classification

Classification is achieved by arranging objects into classes- a class is a collection of objects which share a particular set of properties and no other object has same particular set of properties. In simple words, classification is the grouping together of like objects and their separation from unlike objects. A classification system is called a taxon and the study of a particular taxon, a number of taxons or taxons in general is called as taxonomy, the first really important and successful taxon was the classification of plants and animals [1]. Scientists use classification system to organize information into logically related chunks so that it will be easy to analyze and evaluate. An algorithm that implements classification is known as classifier. The term "classifier" also refers to the mathematical function that maps input data to a category. In machine learning, classification is considered an instance of supervised learning i.e. learning in which training set is a data set whose true classes are known. The corresponding unsupervised learning or clustering is a method in which the classes are inferred from the data on some measure of inherent similarity. Classification is applied in many fields; some of the areas of classification are Library classification, Biological classification, Mathematical classification, Statistical classification.

1.1.1 Binary and multiclass classification

Classification can be considered as two separate problems - binary classification and multiclass classification. In binary classification the members of given set of objects are classified into two groups on the basis of whether they have some property or not for example, medical testing to determine if a person has certain disease or not. On the other hand the multiclass classification involves assigning of an object to one of several classes and it often requires the combined use of multiple binary classifiers.

1.1.2 Statistical Classification

In Statistical classification we identify the sub-population to which the new observation belong, on the basis of a data which contains observations whose subpopulation is known; this data is known as a training set. Statistical approaches are generally based on probability model, which provides a probability of being in each class rather than simply a classification [2].

1.2 Classification Rule

In a given population whose members can be separated into a number of different classes or sets, a classification rule is defined as a procedure in which the each element of the population set is assigned to the class it really belongs. If the every element of the population is assigned to the class it really belongs then the test is perfect and if some error appears then the test is imperfect, and then some statistical analysis must be applied to analyse the classifications. Binary classification is a special kind of classification rule.

1.2.1 Testing Classification Rules

A data set is represented by x and y, where x is each element of the population and y is the class of x, a classification rule is a function which assigns each element to its class. A binary classification is such that the label y can take only a two values.

A classification rule is a function h that can be evaluated for any possible value of x, if the given data is $\{(x_1, y_1), (x_2, y_2)..... (x_n, y_n)\}$ then h(x) will produce a similar classification

$\hat{y} = h(x)$, where $\hat{y}$ is as close as possible to the true group label y. Sometimes the true labels y_i will not necessarily match their approximations $\hat{y}_i = h(x_i)$. In binary classification, the elements that are not correctly classified are named false positives and false negatives.

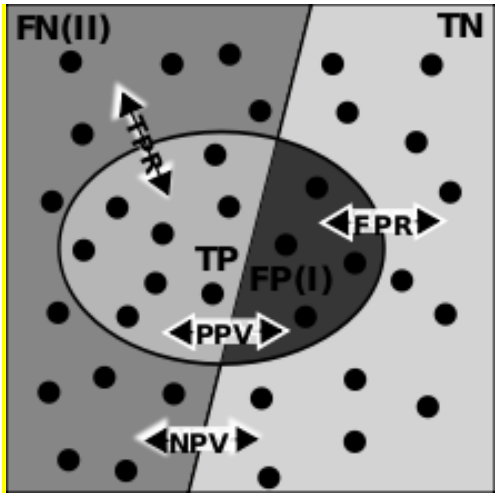

Fig 1.1 Binary-classification labelled

1.2.2 False Positives

False positives occur when a test falsely or incorrectly reports a positive result. For example, a medical test for a disease may return a positive result indicating that patient has a disease even if the patient does not have the disease. We use Bayes theorem to find the probability that a positive result of medical test is in fact a false positive. It is found that if a disease is rare, then the majority of positive results may be false positives, even if the test is accurate.

Let us suppose that the results generated by a test for a disease are as follows:

- The test returns a positive result 99% of the time, or with probability 0.99, if a patient has a disease.
- The test returns a positive result 5% of the time, or with probability 0.05, if a tested patient does not have a disease.

Idealistically, one might think that only 5% of positive test results are false, but that is not right, as we shall see.

Let us suppose that only 0.1% of the population has that disease, such that a randomly selected patient has a 0.001 prior probability of having the disease. To calculate the probability that a positive test result is a false positive we use Bayes theorem.

Let *A* represent the condition in which the patient has the disease, and *B* represent the indication of a positive test result. Then, the probability of actually having the disease given the positive test result is

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|not\ A)P(not\ A)}$$

$$= \frac{0.99 \times 0.001}{0.99 \times 0.001 + 0.05 \times 0.999}$$

$$\approx 0.019$$

Hence the probability of positive result being a false positive is about $1 - 0.019 = 0.98$, or 98%.

Even though the accuracy of test is high, the frequency of disease is so low that the vast majority of patients who test positive do not have the disease. However, the fraction of patients who test positive who do have the disease is 0.019 which is 19 times the fraction of people who have not yet taken the test who have the disease i.e. 0.001. Hence, the test is not useless, and to improve the reliability of result re-testing is required. To reduce the problem of false positives it is required that the test should be very accurate in reporting a negative result when the patient does not have the disease. If the test reported a negative result with the probability of 0.999 in patients without the disease, then

$$P(A|B) = \frac{0.99 \times 0.001}{0.99 \times 0.001 + 0.001 \times 0.999}$$

$$\approx 0.5$$

Such that 1-0.5 = 0.5 (is the probability of false positive).

1.2.3 False negatives

On the other hand, when a test falsely or incorrectly reports a negative result then it is known as false negatives. For example, a medical test for a disease may return a negative result showing that patient does not have a disease even though the patient actually has the disease. We again use Bayes' theorem to calculate the probability of a false negative.

$$P(A|not\ B) = \frac{P(not\ B|A)P(A)}{P(not\ B|A)P(A) + P(not\ B|not\ A)P(not\ A)}$$

$$= \frac{0.01 \times 0.001}{0.01 \times 0.001 + 0.95 \times 0.999}$$

$$\approx 0.0000105$$

The probability that a negative result is false negative is about 0.00105%. False negatives are not a major problem with the test if the disease is rare. But the probability of false negative would be greater if the 60% of the population had the disease. For 60% of population having the disease the probability of a false negative would be

$$P(A|not\ B) = \frac{P(not\ B|A)P(A)}{P(not\ B|A)P(A) + P(not\ B|not\ A)P(not\ A)}$$

$$= \frac{0.01 \times 0.6}{0.01 \times 0.6 + 0.95 \times 0.4}$$

$$\approx 0.0155$$

The probability that a negative result is a false negative rises to 0.0155 or 1.55%.

1.3 Steps of Classification

The various steps to perform the task of classification are as follows:

1.3.1 Acquiring data for Classification

To perform the classification our first step is to acquire or collect the data. In general there are two techniques for the collection of data i.e. 1) Instrumental method for data collection and 2) Survey.

Instrumental method: In Instrumental method data is acquired using some specific instrument for example, to acquire the finger pulse profile we use the data acquisition instrument by BIOPAC. The data acquired is analysed either in time domain or frequency domain.

Survey: Survey based method includes interviews or questionnaires for the collection of data. Questionnaires are the popular mean of collecting the data but difficult to design and usually requires frequent rewrites before an acceptable questionnaire is produced. Questionnaire method has some advantages like it can cover a large number of people and organisations, it can easily be posted e-mailed or faxed and it is relatively cheap. Some of the disadvantages of this method are it has time delay whilst waiting for the responses to be returned, questions have to be relatively simple and we have no control over who completes it. Once a questionnaire is filled by considerable amount of population our data set is ready, but to measure that the questionnaire is reliable for the task of classification we perform reliability test. Reliability is a measure of consistency, if a measure produce consistent results under consistent condition then it is said to have a high reliability. One of the techniques to measure the reliability is Chronbach's alpha test, Chronbach's basic equation for alpha is given as:

$$\alpha = [n/(n-1)\,](1 - \sum V_i \,/\, V_{test}) \qquad - \quad (1)$$

Here, n = number of questions; V_i = variance of scores on each question; V_{test} = total variance of overall scores on the entire test. Alpha varies from 0 to 1, since it is the ratio of two variances. Higher values of alpha are more desirable as it shows more reliability or internal consistency. We can see from the formula of equation 1 that large value of V_{test} gives high value of alpha. High value of alpha is caused by high variance and high variance means we have a wide spread of scores, which means peoples are easier to differentiate. Value of alpha also tells about whether the particular question in questionnaire is good or bad, we check this

by noticing the change in value of alpha if that one question is not on the test. If the value of new alpha comes low it means that the question was good because by deleting that question the overall value of alpha also decreases.[3]

1.3.2 Feature Extraction and Selection

Once the data is acquired the next step is of feature extraction. The task of feature extraction is to get the most relevant information from the original data and represent the information in a lower dimensional space [4]. Sometimes we get our features in simple time domain and sometimes we have to find first derivative and second derivative to extract the features from original data. A good classifier is constructed only if the appropriate set of features is selected. Feature selection is the process of identifying and removing as many irrelevant and redundant features as possible, irrelevant features simply add noise to the data and affect model accuracy [5]. Usually many features depend on one other which often influences the accuracy of classification models. This problem is reduced by constructing new features from the basic feature set and this technique is called as feature construction or transformation [6]. These new features further create more concise and accurate classifiers. It is appropriate to mention here, that several prominent features become available only after transforming the signal from time to frequency domain or to time-frequency domain.

1.3.3 Feasibility Study of Classification

Before building a classifier, many a times the data collected from the predefined (known) classes is checked statistically for the feasibility of classification design. This means that the data from two different known classes has to be statistically different, and not otherwise. This is checked using several statistical methods, of which, the two are briefly discussed below.

Chi-square test: Chi-square test is a statistical test used to check if the difference between expected and observed result is significant, i.e. whether any difference is caused by chance or some other factor is affecting the result according to a specific hypothesis. The formula for calculating Chi-square (χ^2) is:

$$\chi^2 = \sum_{i=1}^{N} (O_i - E_i)^2 / E_i \quad - \quad (2)$$

Here, O_i = observed frequency; E_i = expected frequency [7]. Once we calculate χ^2 using the formula in equation 2, we determine the degrees of freedom and use the chi-square

distribution table to locate the value closest to calculated χ^2 on that degrees of freedom *df* row then move up the column to determine the value of *p*. If the *p* value is $p > 0.05$ then the null hypothesis is accepted, it means that any deviation from expected value is due to chance only. If the $p < 0.05$ then the null hypothesis is rejected, it means that some factor other than chance is operating for the deviation to be so great. Chi-square test has a limitation that it works only with frequency of data that is; we cannot apply this test on the original values of features as it deals only with the frequency of occurrence. To work with original values we use t-test to check the feasibility of classification.

t-test: A t-test is a statistical hypothesis test which determines whether the means of two groups are statistically different from each other. In other words, the t-test compares the actual difference between two means in relation to the variation in data. The formula of t-test is: $t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\left(\frac{{s1}^2}{n_1} + \frac{{s2}^2}{n_2}\right)}}$ - (3)

here, $\overline{x_1}$ = mean of sample 1; $\overline{x_2}$ = mean of sample 2 ; n_1 = number of subjects in sample 1; n_2 = number of subjects in sample 2; ${s_1}^2$ = variance of sample 1; ${s_2}^2$ = variance of sample 2 [8]. After calculating the value of *t* from the formula in equation 3.we find the degrees of freedom *df* i.e. equal to sum of the persons in both groups minus two. Given the value of *t* and *df* we can look up in the standard table of significance to determine whether the t-value is large enough to be significant. If p=0.05 then the null hypothesis is rejected and the sample values do differ from one other, if $p < 0.05$ it shows that the difference in sample values are highly significant.

1.3.4 Training and Testing of Classifier

To train means, is to learn. Therefore training of classifier refers to the learning of classifier system. There are three major learning methods, each corresponding to a particular abstract learning task. These methods are as follows:

Supervised learning: In this type of learning we provide the classifier with a series of sample inputs and comparing the output with the expected response. The training continues until the classifier provides expected response.

Unsupervised learning: If the target output is not known then this type of learning is called as unsupervised learning. This method is more complex and difficult to implement. Unsupervised training method is also called as self learning method because of their ability to carry out self learning. This training process extracts the statistical properties of the training set and groups similar objects into classes. This type of method is adopted in the case of self organizing feature maps, adaptive resonance theory etc.

Reinforcement learning: In this method, the right answer is not presented to the classifier but the teacher is assumed to present. The classifier is only presented with an indication of whether the output answer is right or wrong. The classifier only uses this information to learn and improves its performance. Reinforcement learning is a general approach which is applied where knowledge required to apply supervised learning is not available. Reinforcement learning tries to learn the input-output mapping through trial and error with a view to maximize performance index. The system knows whether the output is correct or not, but does not know the correct output. However, it is usually better to use other methods such as supervised and unsupervised learning, because they are more direct and their analytical bases are well understood.

From all these three learning methods supervised learning is most popular and widely used, unsupervised and reinforcement learning methods are used by artificial neural network classifiers. Both supervised and unsupervised methods have finite training set for the learning of classifier. A training test must contain a list of objects with known classification. Ideally the training set must contain as many examples as possible so that it includes both common and rare type of objects. There are two main concerns before these two methods are used. First, a training set should be constructed for which the true classifications of the objects are known, and second, a set of object parameters must be chosen that are powerful discriminators for classification.

After training the classifier we measure the performance achieved by a learning algorithm. To do this we use a test set consisting of examples with known labels. First we train the classifier using training set then apply it to the test set, and then measure performance by comparing the predicted labels with the true labels [9]. Testing of classifier is also known as evaluation of the classifier. The various techniques to evaluate the classifier are as follows:

Resubstitution: In this technique we first use all the available data to design a classifier the we use same data to test the classifier. As in this technique we use all the available data for design therefore we call the classifier an "optimal" classifier. This technique is easy and fast. Using the same data for training and testing gives an optimistically biased estimate of the error rate, but variance is relatively smaller as all data is used for training. Since it suffers from "testing on the training data," resubstitution is not recommended unless the training set is very large.

Holdout (data partition): In this method data is separated into two groups of training and testing. The size of training set is about twice of testing set. This classifier is sub-optimal in the sense that it uses only part of available data for training. Testing suffers from small a small test sample. Unlike resubstitution method its result estimate of error rate is unbiased, but has large variance which implies uncertainty. It is also easy and fast in performance.

Cross-validation: This method is generalization of holdout method. In this technique N total samples are divided into *m* groups of equal size. Then *m* different classifiers are trained each using *m* –1 groups, holding out each of the groups. For each of the *m* classifiers, the group left out is tested. Then the *m* test results are averaged. All samples get used for both training and testing. The result is unbiased and with minimum variance. This method gives good results when a large number of samples are available.

Jack-knife (Leave-one-out): This technique is a limiting case of cross-validation. Where *m* = N and N different classifiers are trained each using N–1 samples. For each of the N classifiers, the one left out sample is tested. Then the N test results are averaged. Classifiers are very close to optimal as all samples are used for testing. It is the best method for estimating performance. The result is unbiased and with minimum variance. If a fast leave-one-out algorithm is available (to estimate necessary parameters using an update scheme: e.g., mean, covariance and its inverse and determinant), this method is fairly fast and useful for Bayes quadratic and k-nearest neighbour. But, if no fast algorithm is available, then it is very slow, like neural networks.

Bootstrap: This method has various forms, a common one is: A training set is generated by randomly selecting N samples using replacement (i.e., samples can be selected more than once). The samples which are not selected for training are used for testing. The process is repeated many times, for e.g., 200. The results are averaged to give final estimate of the error

rate. The classifiers produced are never optimal and the resulting estimate is unbiased. This method is very computationally intensive which makes it slow. This method is very good for use when only a few samples are available, say less than 30. [10]

In designing of classifier if we have more training data then it gives better generalization and if we have more test data then it gives better estimate for classification error probability.

1.4 Classification Techniques

There are various classification techniques ranging from simple techniques such as Rule based and Nearest-Neighbour classifiers, to more advanced techniques such as Support Vector Machine.

Rule based classifier: In this technique a collection of "if..then.." rules are used for classifying records. Each classification rule is expressed in the following way:

r_i : $(Condition_i) \longrightarrow y_i$

The left-hand side of the rule is known as the rule antecedent or precondition.

It contains a conjunction of attribute tests:

$Condition_i = (A_1\ op\ v_1) \wedge (A_2\ op\ v_2) \wedge ..... (A_k\ op\ v_k)$,

Where *op* is a logical operator chosen from the set {=, ≠,< ,>, ≤ , ≥ } and (A_j , v_j) is an attribute value pair. Each attribute test ($A_j\ op\ v_j$) is called as conjunct. The right-hand side of the rule which contains the predicted class y_i is known as the rule consequent. A rule *r* covers a record *x* if the precondition of *r* matches the attributes of *x* and *r* is said to be fired and triggered whenever it covers a given record. A rule-based classifier classifies a test record on the basis of rule triggered by the record. To build a rule-based classifier, we need to extract a set of rules that represents key relationships between the attributes of a data set and the class label. There are mainly two broad classes of methods for extracting classification rules:

(1) direct methods, which extract classification rules directly from data, and
(2) indirect methods, which extract classification rules from other classification models, such as decision trees and artificial neural networks.[11]

Nearest- Neighbour classifier*:* In this classification technique, we simply find all the training examples that are relatively similar to the attributes of test example. These examples are known as nearest neighbours and can be used to determine the class label of the test example. A nearest neighbour classifier represents each example as a data point in a d-dimensional space, where d is the number of attributes. The justification for using nearest neighbour is best explained by following saying: "*If it walks like a duck, quacks like a duck, and looks like a duck, then it's probably a duck*" [12]. Or we can say, since the neighbour is nearby, it is likely to be similar to the object being classified and so is likely to be the same class as that object.

Naive Bayes classifier: This classifier assumes that the presence (or absence) of one attribute of class is unrelated to the presence (or absence) of any other attribute, given the class label *y*.

The conditional independence assumption can be formally stated as follows:

$P(X|Y = y) = \prod_{i=1}^{d} P(X_i | Y = y)$,

Where each attribute set $X = \{X_1, X_2, X_3, \ldots\ldots\ldots\ldots, X_d\}$ consists of *d* attributes.

With the conditional independence assumption, instead of computing the class conditional probability for every combination of X, we only have to estimate the conditional probability of each X_i , given Y [13]. For example, a fruit may be considered to be an orange if it is orange in colour, round in shape, and about $4''$ in diameter. Even if these attributes depend on each other or upon the existence of the other attribute, a naive Bayes classifier considers all of these properties to independently contribute to the probability that this fruit is an orange.

Artificial Neural Network: An Artificial Neural Network (ANN) is a mathematical model that is inspired by attempts to simulate biological neural systems. Analogous to human brain structure an ANN is composed of an interconnected assembly of nodes and directed links. A neural network consists of an interconnected group of artificial neurons, and it processes information using connectionist approach to computation. An ANN is an adaptive system during learning phase it changes its structure based on external or internal information that flows through the network. An ANN is typically defined by three types of parameter:

1. The interconnection pattern between different layers of neurons - single layer neurons, multilayer neurons, recurrent networks.

2. The learning process for updating the weights of interconnections- supervised learning, unsupervised learning, reinforcement learning
3. The activation function that converts a neuron's weighted input to its output activation- linear(or ramp), threshold, sigmoid.

Support Vector Machine (SVM): This technique has its roots in statistical learning theory and it has shown very promising results in many practical applications, from handwritten digit recognition to text categorization. It also works efficiently with high dimensional data and avoids the problems associated with dimensionality. An SVM training algorithm builds a model that assigns new examples into one category or the other. An SVM model maps the examples as points in space such that the examples of the separate categories are divided by a clear gap that is as wide as possible as shown in Fig.1.

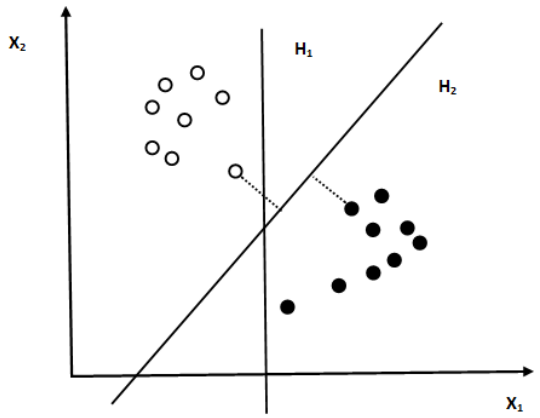

Fig.1.2 SVM maximum Separating Hyper plane (H_2) with margin

New examples are then mapped into that same space and their category is predicted, based on which side of the gap they fall on. A support vector machine constructs a hyper-plane or set of hyper-planes in a high-dimensional space and a good separation is achieved by the hyper-plane that has the largest distance to the nearest training point of any class, these nearest points are called as support vectors.

1.4.1 Advantages and Disadvantages of various techniques and their application

Rule based classifier: Rule based classifier is easy to interpret and easy to generate, it can classify new instance rapidly, but defining rules can be tedious for large data set with many categories because as data set grows we have to write more rules correspondingly. The rule based classifier is best suited for the area where number of rules required for classification is small and accuracy is sufficiently high. For example, vertebrate classification, malnutrition detection in children (Xu Dezhi et al.). [14]

Nearest-Neighbour classifier: The nearest neighbour algorithm is intuitive and easy to understand which facilitates implementation and modification. It provides good generalisation accuracy on many domains. The most serious shortcoming of nearest neighbour technique is that they are very sensitive to presence of irrelevant parameters, this technique requires large storage because it has to store all the data and the method is also slow during instance classification because all the training instances have to be visited. Nearest-Neighbour classifier has its application in the area of pattern recognition (particularly for optical character recognition), Spell checking, DNA sequencing.[15]

Naive Bayes classifier: This classifier is fast to train and evaluate, it gives very good results in real world problems. Naive Bayes is robust to isolated noise points because these points are averaged out when estimating conditional probabilities from data. The disadvantage of Bayesian classifier is that the assumption of class conditional independence usually does not hold which degrade the performance of classifier therefore it is not suitable for more complex classification problems where features are usually correlated. Bayesian classifiers are very well suited for problems involving normal distributions which are very common in real world problems, most email clients use Naive Bayes classifiers for filtering out spam emails.[16]

Artificial Neural Network (ANN): The biggest advantage of ANN algorithm is that they are general: they can handle problems with many parameters and produces remarkable results in complex domain. ANN is suitable for both continuous and discrete data, it works efficiently even in the presence of redundant features because the weights are learned automatically during the training step and the testing is very fast. The disadvantage of neural networks is that they are very slow especially in training phase and it is very difficult to determine how the network is making its decision. Neural networks are also sensitive to the presence of noise

in training data. The application of ANN includes image classification, speech recognition, weather forecasting, cancer detection, prediction of stock market performance.[17]

Support Vector Machine (SVM): This algorithm captures the inherent characteristic of the data better than ANN, SVM is able to handle large feature space as its complexity does not depends upon the dimensionality of the feature space and SVM is the most powerful non linear classifier. The disadvantage of SVM is that it is sensitive to noise and computationally demanding to train and run. In SVM the choice of kernel function and the tuning of parameters have to be done manually which greatly impacts the result. The applications of SVM classification technique are text categorization, bioinformatics (protein classification), handwritten character recognition.[18]

1.5 Issues

There are many issues of concern to the would- be classifier. The lists of a few of the issues are given below:

Accuracy: The accuracy of classifier is an important concern. The reliability if the classification rule is usually reported by the proportion of correct classifications. Sometimes some errors are more serious than others therefore it is important to control the error rate.

Speed: The speed of the classifier is also a major issue. A classifier that is 90% accurate may be preferred over one that is 95% accurate if it is 100 times faster in testing. For example, such considerations would be important for the automatic reading of postal codes or automatic fault detection of items on a production line.

Comprehensibility: In case of human operator to apply the classification procedure there is chance that the operator will make mistake in applying the classification rule if procedure is not easy to understand. Another important aspect is that human operator must believe the system. An oft-quoted example is the Three-Mile Island case, where the automatic devices correctly recommended a shutdown, but this recommendation was not acted upon by the human operators who did not believe that the recommendation was well founded.

Time to Learn: In a rapidly changing environment, it is necessary to learn a classification rule quickly, or make adjustments to an existing rule in real time. "Quickly" might also imply that we need only a small number of observations to establish our rule. [2]

Chapter-2

CONCEPT REVIEW

The thesis work is based on concept related to Ayurveda, Nadi parikshan, Plethysmograph and t-test. A review of the basic idea for those is presented here.

2.1 Ayurveda

Ayurveda is the oldest surviving complete medical system in the world. Ayurveda is made up of two Sanskrit words: *Ayus* which means life and *Veda* which means the knowledge. Ayurveda is thus the Science of life, knowledge about life or a sensible way of living based on knowledge. It has its origins go back to nearly 5000 years, when it was practiced by the spiritual rishis who laid the foundation of the Vedic civilisation in India, by organising the fundamentals of life into proper systems. Therefore, main source of knowledge in this field are present in the *Vedas* and more specifically the fourth of the series, namely *Atharvaveda* that dates back to around 1000 BC. The most famous foundational works of Ayurveda are *Charaka Samhita* and the *Sushruta Samhita* which concentrate on internal medicine and surgery respectively.

Ayurveda is based on Sankhaya's philosophy of creation and manifestation. Which professes that behind all creation there is a state of pure existence or awareness, which is beyond time and space, has no beginning or end, and no qualities. Within pure existence, the desire to experience itself arises, which results in disequilibrium and causes the manifestation of the fundamental physical energy. As Matter and energy are closely related such that when energy takes form, we tend to think of it only in terms of matter. And when this energy is much modified, it ultimately leads to the manifestation of our familiar mental and physical worlds. It also gives rise to cosmic consciousness, which is the universal order which affect strongly our whole life. Individual intelligence, as different from the everyday intellectual mind, is derived from and is part of this consciousness. It is the inner wisdom or the part of individuality that remains unmoved by the demands of daily life, or by *Ahamkara*, the sense of `I-ness'. A Sanskrit word Ahamkara is a concept that is not quite understood by everyone

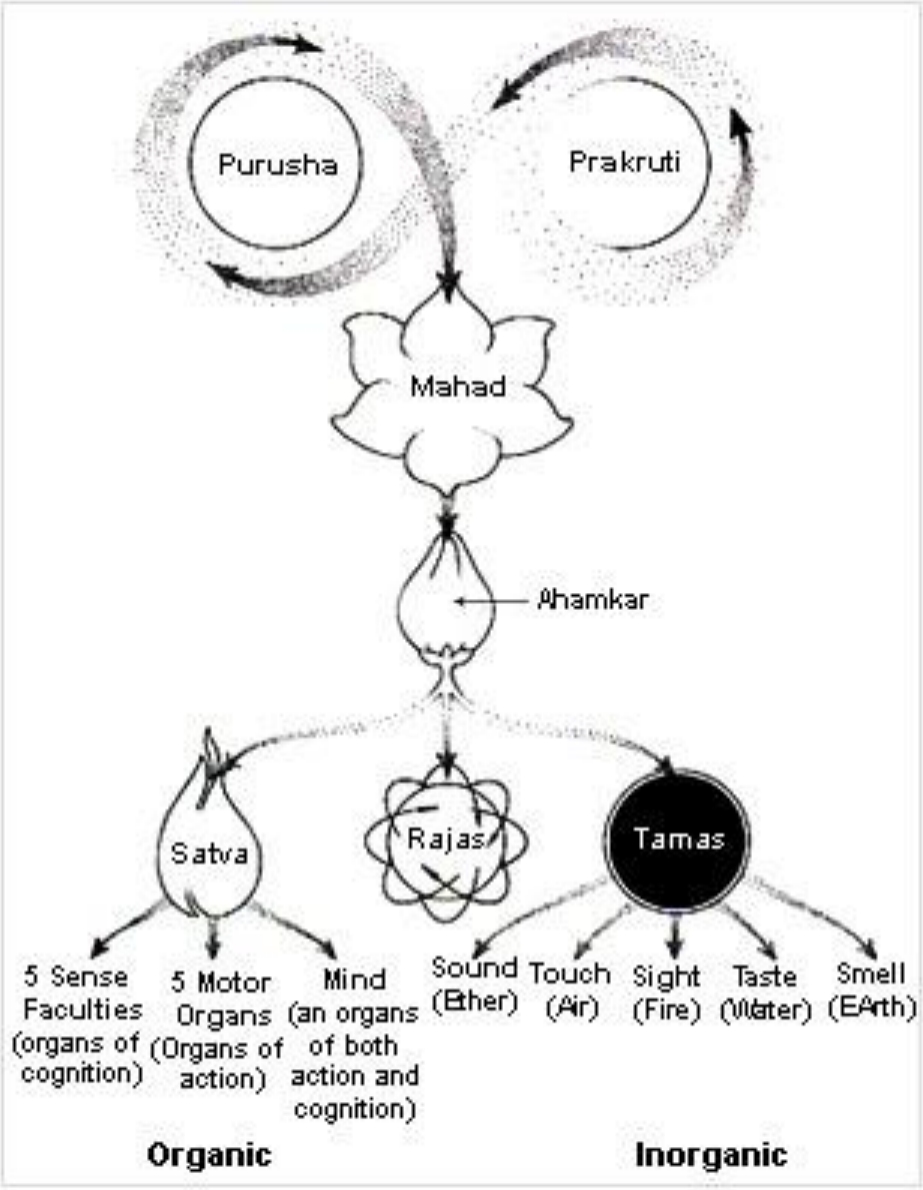

Fig. 2.1 Sankhya Philosophy of Creation

as it is often equated to `ego'. But it embraces much more than just that, it tells that 'I' am not separate from the universal consciousness, but 'I' has an identity that differentiates and defines the boundaries of `me'. All creations therefore have Ahamkara, not just human beings. A two-fold creation arises from Ahamkara, The first is Satwa, the subjective world, which is able to observe and manipulate matter. It comprises the subtle body (the mind), the capacity of the five sense organs to hear, feel, see, taste and smell, and for the five organs of action to speak, grasp, move, procreate and excrete. The second is Tamas, the objective world of the five elements of sound, touch, vision, taste and smell – the five subtle elements that give rise to the dense elements of ether or space, air, fire, water and the earth – from which all

matter of the physical world is derived. And it is Rajas, the force or the energy of movement, which brings together parts of these two worlds. Satwa, Rajas and Tamas also known as Trigunas are three essential energies of mind. Satwa is characterised by lightness, consciousness, pleasure and clarity, it is pure, free from disease and cannot be disturbed in any way. It activates the senses and is responsible for the perception of knowledge. And Rajas is the most active of the gunas its characteristics are motion and stimulation. All desires, wishes, ambitions and fickle-mindedness are a result of the Rajas guna. The last Tamas is characterised by heaviness and resistance. It produces disturbances in the process of perception and activities of the mind. It causes delusion, false knowledge, laziness, apathy, sleep and drowsiness. [19]

2.1.1 Tridoshas

According to Ayurveda our body is made up of five primary elements: air, water, earth, fire and ether (space) together called the *Panchamabhutas* which arise from Tamas. Just like nature we too have these five elements in us, and they have ability to combine and to create various physiological functions.

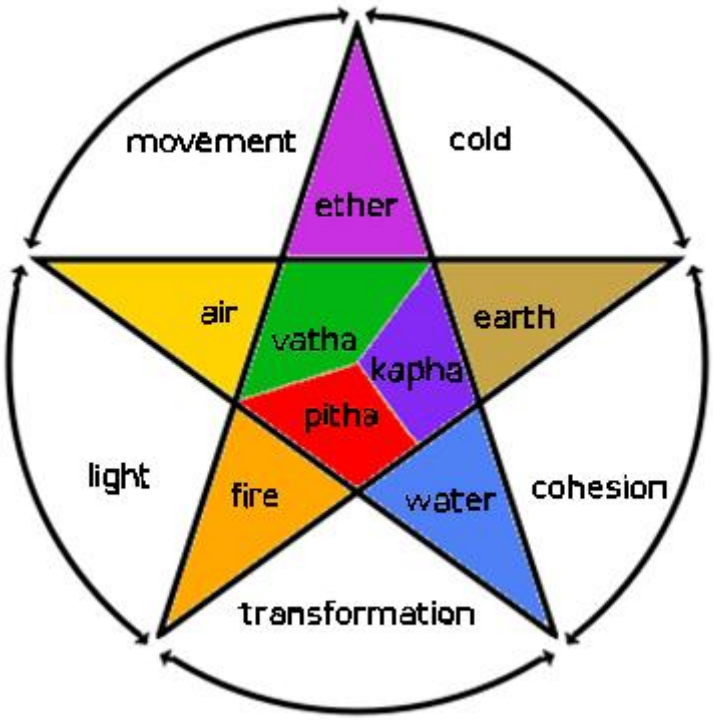

Fig 2.2 The three doshas and the five elements from which they are composed.

Ether and air combine to form what is known in Ayurveda as the *Vata* dosha, relating to physical movements and sensory perception; fire and water are the elements that combine to form the *Pitta* dosha, which is connected with metabolism and functioning of the digestive process; and finally water and earth elements combine to form the *Kapha* dosha, relates to moisture in the body tissues and sensory organs. Therefore, every physical characteristic, mental capacity and the emotional tendency of a human being can be explained in terms of the tridoshas. When the doshas are in balance i.e. in state of equilibrium, we remain healthy. But if there is imbalance in the functioning of any one or two of the tridoshas the illness in our body arises.

2.1.2 Nadi Pariksha (pulse diagnosis)

In traditional medical language Nadi means channels. These channels exist within the body in the form of blood vessels, nerves, lymphatic channels, nodes etc. According to Ayurveda each cell in our body possesses its own intelligence and communication of this intelligence takes place in the form of vibrations through these channels [20]. Nadi gets influenced by various affects that influence the physiology of the body, especially the three dosha i.e. Vata, Pitta and Kapha. In medical terms Nadi refers to the radial artery. The motion of Nadi does not only indicates the rate or rhythm of the pulse, the size and condition of the arterial wall or the volume of blood flows through the artery or the force with which the blood flows, it also indicates the imbalance of the three doshas. The Nadi Pariksha is chiefly performed to ascertain the imbalance of dosha or doshas.

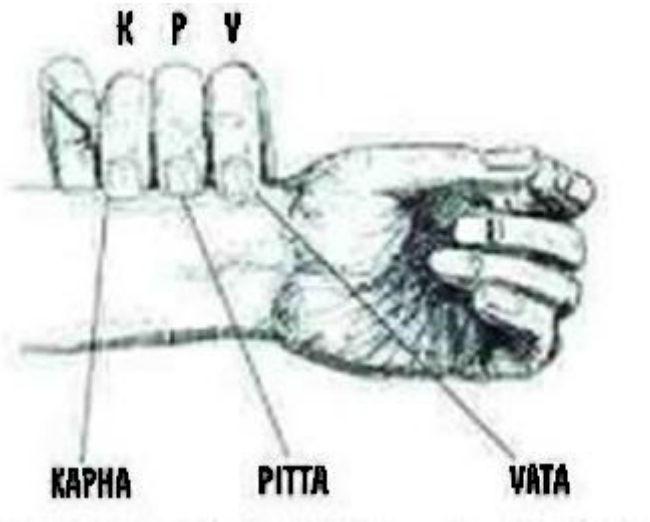

Fig 2.3 Nadi Pariksha (traditional method)

In Nadi Pariksha to examine the pulse three fingers are used because the nature of mind causes great difficulty in assessing the three dosha i.e. Vata, Pitta and Kapha with one finger. Evaluation of pulse from three fingers has been converted into a standard rule and that is as follows: Vata is established by the tip of the index finger of right-hand, placed on the radial artery next to the root of the thumb, the Pitta pulse should be studied by the touch of the tip of the middle finger placed next to it and the Kapha pulse by touch of the tip of the ring finger placed next to the middle finger on the artery.

2.2 Photoplethysmography

Photoplethysmography (PPG) is a non invasive method for the measurement of arterial and venous blood volume changes at a peripheral site where the blood vessels are close to the skin, for example, at the finger and the earlobe [21]. It provides the valuable information about the cardiovascular system. The basic PPG technology requires to optoelectronic components: first, a light source to illuminate the tissue (e.g. skin) and second, a photo detector to measure the small variations in light intensity associated with changes in perfusion in the catchment volume [22]. It operates at a red or near infrared wavelength because there we have the strongest modulation of the signal due to light absorption in the haemoglobin in the blood. The most recognised waveform feature is the peripheral pulse, which is synchronized to each heartbeat. The advantages of PPG are: First, it is non-invasive and the PPG signal is strong and robust. Second, the electronic circuitry is simple and small. Third, it is safe because there is no electrical contact to the human body and can also be used in water.

Fig 2.4 Photoplethysmograph Sensor

The two PPG operational configurations are: Transmission (trans-illumination) mode operation where the fingertip is placed between the source and detector, as shown in figure 2.5

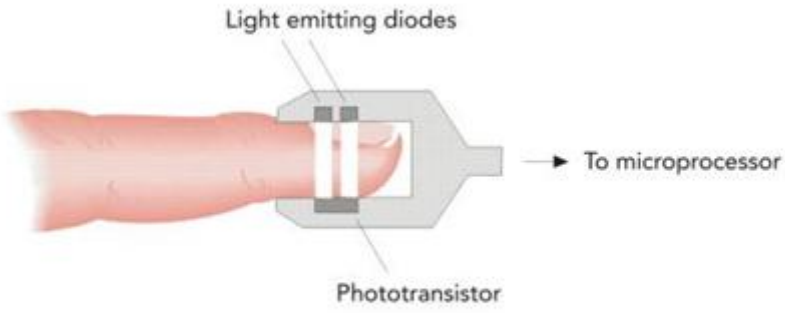

Fig 2.5 PPG sensor in transmission mode.

and reflection (adjacent) mode operation where the LED and detector are placed side-by-side. The detected optical radiation waveform consists of a pulsatile ("AC") component which results from cardiac synchronous changes in the blood volume with each heart beat. This AC component is superimposed on a slowly varying ("DC") baseline with various lower frequency components which results from respiration, sympathetic nervous system activity and thermoregulation [23].

2.2.1 Features of Finger PPG

The Pulse Wave (PW) is a complex physiological phenomenon observed and detected in blood circulation. In the course of heart systole a certain amount of blood is ejected and it is moved into the arteries because of transformation between kinetic and potential energy of each segment of ejected blood. On each artery or venous section affected by a pulse wave, three coherent phenomena are observed: blood flow (flow pulse), the increase of blood pressure (pressure pulse) and extension of transverse profile (profile or volume pulse).

The typical volume pulse wave obtained from finger PPG is shown in Fig2.6.

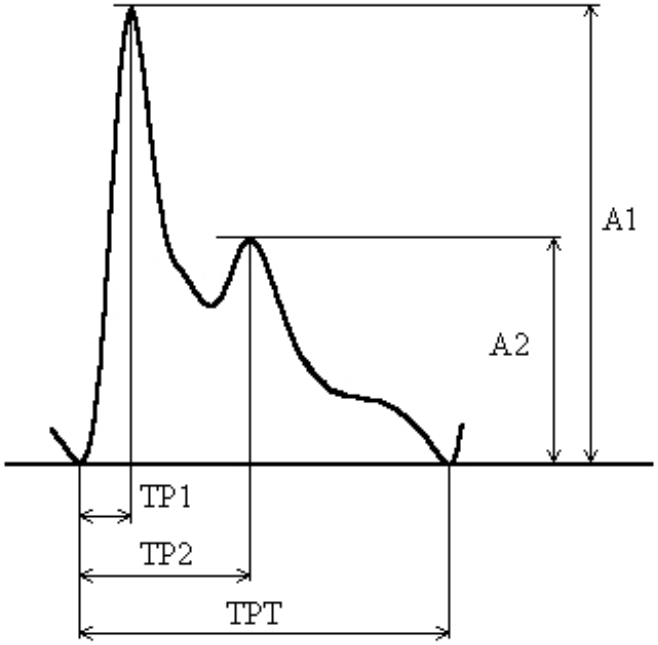

Fig 2.6: Typical volume or pressure pulse wave of peripheral artery

On the time course of pulse wave there are following proportions which can be measured: crest time (TP1), dicrotic wave time (TP2), total pulse duration (TPT), systolic amplitude (A1), dicrotic wave amplitude (A2). From these proportions following parameters or we can say features can be derived: relative crest time (RCT) = TP1/TPT, relative dicrotic wave time (DWT) = TP2/TPT and relative dicrotic wave amplitude (DWA) = A2/A1, all these parameters are dimensionless [24].

2.3 t-test

A t-test is a statistical hypothesis test which determines whether the means of two groups are statistically different from each other. In other words, the t-test compares the actual difference between two means in relation to the variation in data. The t-statistic was introduced in 1908 by William Sealy Gosset, a chemist working for the Guinness brewery in Dublin, Ireland."Student" was his pen name. Gosset devised the t-test as a cheap way to monitor the quality of stout. He published the test in Biometrika in 1908, but was forced to use a pen name by his employer, who regarded the fact that they were using statistics as a trade secret. [25]

To understand the concept of t-test an example is given below where three situations are considered in which the difference between the means is the same in all three yet the mean of each is statistically different. The three situations are shown in figure 2.7

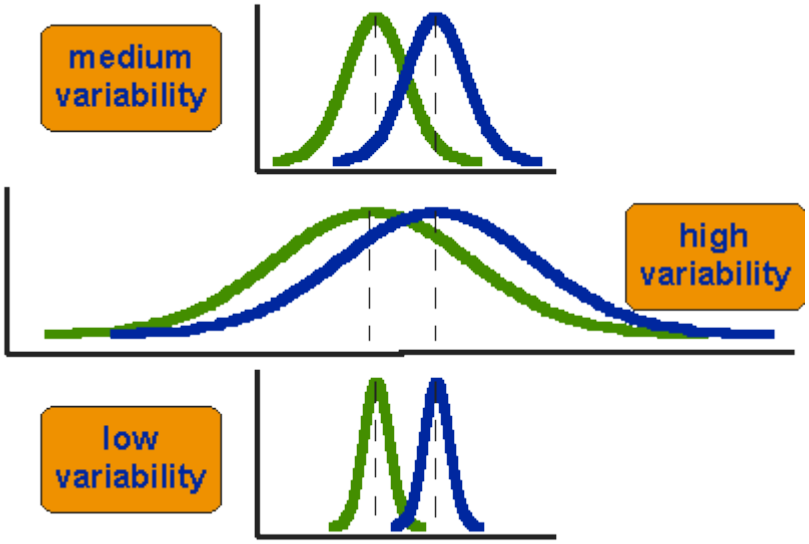

Fig 2.7 Three scenarios for differences between means.

In these three situations the difference between the means is same but the three situations don't look same, they each represent different story. The example on top shows a case with moderate variability of scores within each group. The situation in middle shows the high variability case and the third situation shows the case with low variability. In the last case the two groups appear most different because there is little overlap between the two bell-shaped curves. In the high variability case, the group difference appears least striking because the two bell-shaped curves overlap so much.

This leads to a very important conclusion that when we are looking at the differences between scores for two groups, we have to judge the difference between their means relative to the spread or variability of their scores. And this is exactly the work of t-test, to compare the actual difference between two means in relation to the variation in data.

The various t-tests that are most frequently used are:

- A one-sample location test to find whether the mean of a normally distributed population has a value specified in a null hypothesis.
- A two sample location test of the null hypothesis that the means of two normally distributed populations are equal. These t-test are called as Student's t-test as this test is used only when the variances of the two populations are assumed to be equal; the form of the test where this assumption is not used is called as Welch's t-test. These tests are often referred to as "unpaired" or "independent samples" t-tests, as they are typically applied when the statistical units underlying the two samples being compared are non-overlapping.
- A test of the null hypothesis that the difference between two responses measured on the same statistical unit has a mean value of zero. For example, suppose we measure the size of a cancer patient's tumor before and after a treatment. If the treatment is effective, we expect the tumor size for many of the patients to be smaller following the treatment. This is often referred to as the "paired" or "repeated measures" t-test.
- A test of whether the slope of a regression line differs significantly from 0.

2.3.1 Unpaired and paired two-sample t-tests

Unpaired: The unpaired, or "independent samples" t-test is used when two separate sets of independent and identically distributed samples are obtained and one from each of the two populations is compared. For example, suppose the effect of a medical treatment is evaluated on the 100 subjects from which 50 subjects are selected for treatment group and another 50 subjects are selected for control group randomly. In this case, the two independent samples are constructed and the unpaired form of the t-test should be performed.

Paired: Paired or "Dependent samples" t-tests typically consist of a sample of matched pairs of similar units, or one group of units that has been tested twice which is known as a "repeated measures" t-test. A typical example of the repeated measures t-test is; where subjects are tested prior to a treatment, say for high blood pressure, and the same subjects are tested again after treatment with a blood-pressure lowering medication.

2.3.2 Statistical Analysis of t-test

The formula of the t-test from equation 3:

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\left(\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}\right)}}$$

The formula for the t-test is ratio. The top part of ratio is just the difference between the two means or averages. The bottom part is a measure of variability or dispersion of scores. This formula is easy to understand with an example of signal-to-noise metaphor in research: the difference between the means is the signal that our program introduced into the data; the bottom part of the formula is a measure of variability that is basically noise that may make it harder to see the group difference. The top part of the formula is easy to compute, the bottom part is called as 'standard error of the difference'. To compute the bottom part the variances of each group is taken and divide it by the number of people in that group, then we add these values and take their square root.

Once the t-value is computed, we have to look it up in a table of significance to test whether the ratio is large enough to say that the difference between the groups is not likely to have by 'chance'. To test the significance, we need to set an alpha level which is in most social research as the "rule of thumb" is set at 0.05. This means that five times out of a hundred we would find a statistically significant difference between the means even if there was none (i.e., by "chance"). We also have to find the value of the degree of freedom (*df*) for the test. In the t-test, the degree of freedom is the sum of the persons in both groups minus 2. Given the alpha level, the *df*, and the t-value, we can look the t-value up in a standard table of significance to determine whether the t-value is large enough to be significant. If it is, we can conclude that the difference between the means for the two groups is statistically different.

Chapter-3

LITERATURE REVIEW

3.1 Standard Statistical Classifications: Basic Principles

In February 1999, Eivind Hoffman from Bureau of Statistics, New York discussed the best practices for the development, use, maintenance and revision of International Standard Statistical Classifications. The paper describes that the classification is the task of grouping and organizing information meaningfully and systematically into a standard format that is useful for determining the similarity of ideas, events, objects or person and a statistical classification is a classification having a set of discrete categories, which may be assigned to a specific variable registered in a statistical survey or in an administrative file, and used in the production and presentation of statistics [26].

3.2 Optimization and Interpretation of Rule – Based Classifiers

Machine learning methods are frequently used to create rule- based classifiers. Rule- based classifiers are useful only if rules are reliable, accurate, stable and sufficiently simple to be understood. In 2007 W.Duch, N. Janowski, K. Grabczewski and R. Adamczak from Department of Computer methods of Nicholas Copernicus University, Poland have talked about the optimization and interpretation of sets of rules. The method is equivalent to a specific fuzzification of crisp membership functions, equivalent to an assumption of uncertainties in the inputs. Analysis of the change of probabilities of classification in response to the change in uncertainties allows to estimate confidence in the performance of a rule-based system. If the confidence is low a more detailed analysis of the influence of each feature on classification probability is started [27].

3.3 A Survey of Image Classification Methods and Techniques for Improving Classification Performance

In this March 2007 review published in International Journal of Remote Sensing, Volume 28, issue 5, D. LU and Q. WENG discussed current practices, problems, and prospects of image classification. They emphasized on major advanced classification approaches and the

techniques used for improving classification accuracy. In addition they also described the issues affecting classification performance, Effective use of multiple features of remotely sensed data and the selection of a suitable classification method are especially significant for improving classification accuracy. The paper concludes that the success of an image classification depends on many factors. The availability of high-quality remotely sensed imagery and ancillary data, the design of a proper classification procedure, and the analyst's skills and experiences. For a particular study, it is often difficult to identify the best classifier due to the lack of a guideline for selection and the availability of suitable classification algorithms to hand. For this, the combination of different classification approaches has shown to be helpful for improvement of classification accuracy [28].

3.4 Multi-label Classification: An Overview

In 2007 G. Tsoumakas, I. Katakis from the Department of Informatics, Aristotle University of Thessaloniki, Greece presented a paper on multi-label classification methods. This paper introduced the task of multi-label classification, the paper introduced the problem, gave an organized presentation of the methods that exists in the literature and provided comparative experimental results for some of these methods [29].

3.5 Supervised Machine Learning: A review of Classification Techniques

In July 16, 2007, S. B. Kotsiantis from the department of Computer Science and Technology of University of Peloponesse, Greece presented a paper which explained the supervised learning based classifier techniques in detail. It also discussed the concept of feature selection for building the classifier and the techniques used to calculate the classifier's accuracy. The paper describes the best-known supervised techniques in relative detail. After a better understanding of the strengths and limitations of each method, the possibility of integrating two or more algorithms together to solve a problem was proposed. The objective for that was to utilize the strengths of one method to complement the weaknesses of another [30].

3.6 Making Sense of Chronbach's alpha

This paper was published by Mohsen Tavalok and Reg Dennick in 2011. In this paper they described the meaning of Chronbach's alpha, use of Chronbach's alpha and focused on it as index of reliability. They discussed that high quality tests are important to evaluate the

reliability of data supplied in an examination or a research study. Alpha is a commonly employed index of test reliability. Alpha is affected by the test length and dimensionality. A longer test increases the reliability of a test regardless of whether the test is homogenous or not. A high value of alpha (>0.90) may suggest redundancies and show that the test length should be shortened [31].

3.7 Healthcare and Disease Management in Ayurveda

The disharmony in mental doshas (satogun, rajogun, and tamogun) and body doshas (Vata, Pitta, and Kapha) are the major cause of illness, the goal illness management in Ayurveda is to bring back harmony among the doshas. In 2001, L Mishra, BB Singh, S Dagenais from Southern California University of Health Sciences, USA described in their paper that the management includes clinical examination, diagnosis, and dietry and lifestyle interventions and treatment. The clinical examination consists of Astha Sthana Pariksha (8-point diagnosis: pulse-diagnosis, urine, stool, tongue, voice and body sound, eye, skin, and total body appearance examinations) and examination of the digestive system and the patient's physical strength. The treatment consists of cleansing (Panchkarma), palliation (improve digestion, remove toxic waste, fasting, observe thirst, exercise, sunbathing, and meditation), mental nurturing, and spiritual healing depending on the disturbed doshas and the patient's constitution. The preferred use of bhasms and herbal formulas over the respective metallic salts or the single herbs is discussed. This paper suggested the great potential for integration of Ayurvedic therapies into the healthcare system in the United States [32].

3.8 Contour analysis of the photoplethysmographic pulse measured at the finger

In this paper published in 2006 the application of volume pulse in pulse wave analysis is discussed. This review describes the background to digital volume pulse contour analysis, how the technique relates to contour analysis of the pressure pulse. Volume pulse may conveniently be acquired optically from a finger (digital volume pulse). Although less widely used, this technique deserves further consideration because of its simplicity and ease of use. As with the pressure pulse, the contour of the digital volume pulse is sensitive to changes in arterial tone induced by vasoactive drugs and is influenced by ageing and large artery stiffness. Measurements taken directly from the digital volume pulse or from its second derivative can be used to assess these properties [33].

3.9 Photoplethysmography and its application in clinical physiological measurement

In this 2007 topical review, John Allen from Regional Medical Physics Department of Freeman Hospital, UK describes the basics of Photoplethysmography, PPG is a simple and low-cost technique that can be used to detect blood volume changes in the microvascular bed of tissue. The review also demonstrated PPG's great potential to be used in wide range of clinical measurements. The review focuses on the early and recent history of PPG, pulse wave analysis, instrumentation, clinical physiological monitoring, vascular assessment and autonomic functions [22].

3.10 Parameters Describing the Pulse Wave

A review article presented by D. Korpas, J. Hálek, L. Doležal from the Department of Medical Biophysics, Faculty of Medicine, Palacký University Olomouc, Czech Republic respectively in July 2008 describes the principle of pulse wave measurement, parameters of pulse wave measurement and current clinical practices. They discussed the Pulse wave, types of pulse waves, measurement of pulse wave in time domain. It shows the potential utilization and inaccuracy of clinical measurements. It establishes that the Pulse waveform is a real physiological signal. The measurement is sensitive to body motion as influenced by other physiological rhythms. Despite limitations, this technology represents a promising non-invasive tool for reflecting the status of cardiovascular system both experimental and in a clinical setting [24].

3.11 The shape and dimensions of Photoplethysmographic Pulse Waves: A Measurement Repeatability Study

Zbignevs Marcinkevics, Signe Kusnere, Juris Imants Aivars, Uldis Rubins, Aram Hussain Zehtabi, in 2009, recorded two PPG signals (from the radial artery and from the finger) several times in resting conditions to evaluate the individual repeatability of PPG wave shape. Arbitrary amplitudes of incisura and maximum of dicrotic notch and time from foot to anacrotic maximum had the greatest repeatability among PPG wave shape parameters in both radial artery and finger. Maximum of the first derivative in the anacrotic phase in the PPG waveform showed the highest variability. A possible explanation for the existing individual variability of PPG parameters might be temporal and spatial summation of fluctuations of

heart cycle length and changes of peripheral resistance in magistral artery and microcirculatory vessels [34].

3.12 E- Health Analysis Element for Supporting Therapeutic through Ancient Indian Medical Science

In 2010, a paper by G.M. Kadhar Nawaz *et al* published in International Journal of Computer Applications, Volume 1- No.17 proposed a method of creating a wearable E- Health analysis element, which will support the medical science of swayam Nadi Parikshan. They mentioned that they are working on the development of an element which could greatly support the medical treatment using Nadi Pariksha. The convenient, inexpensive, painless, and non-invasive E- Health analysis element extracts the imbalances of Tridosha, which in turn identifies the presence and location of disorders in a patient's body [35].

3.13 Correlation Studies of Finger Pulse Profiles for Detecting Ayurvedic Doshas

Dr. Mandeep Singh and Spiti Gupta in this paper in 2011 have explored the detection of tridoshas using modern PhotoPlethysmoGraphy (PPG). As per the ancient science of Ayurveda human beings suffer primarily on account of imbalance in the three basic human constituents called doshas. These doshas are normally detected by the skilled masters simply by pressing the radial artery in the wrist with their three fingers. In this study the pulse profiles of all 10 fingers in 7 healthy subjects is acquired using Biopac MP150 data acquisition system. In all 70 cases without exception the autocorrelation for a given finger of subject is always higher than correlation with corresponding finger of any other subject [36].

Chapter-4

PROBLEM FORMULATION

The task of classification occurs in a wide range of human activity. We know that the Ayurveda is the Science of life which professes that our body is made up of five primary elements: air, water, earth, fire and ether (space). These five elements manifest in human body as tridoshas namely Vata, Pitta and Kapha. These five elements have ability to combine and to create various physiological functions from which fire and water elements combine to form a Pitta dosha, which is connected with metabolism and functioning of the digestive process. Most of the illness in our body arises from an imbalance in the functioning of any one or two of the tridoshas. Pitta naturally increases during the middle of life, the day, and the middle of digestion therefore the significant rise in the Pitta dosha is mainly in daytime i.e. in afternoon and after lunch [37].

In Ayurveda, the ancient science of medicine, it is believed that the level of tridoshas in our body can be predicted by Nadi Parikshan i.e. Pulse diagnosis. In modern times, finger pulse or the peripheral pulse which is the extension of radial pulse can be acquired using Photoplethysmograph (PPG). A Recent research proposes that PPG can be utilized as an alternative to Nadi Parikshan. Therefore the pulse waveform acquired by Photoplethysmograph can be used to detect the level of Pitta in body before and after lunch, but we have to check that the change in the Pitta level after the meal is significant enough that it will be used for the task of classification. The aim of this thesis is to check the feasibility of classification of high Pitta using features extracted from finger Photoplethsmogram.

Chapter-5

METHODOLOGY

5.1 Data Collection for Classification

To perform the task of classification the first step is to acquire or collect the data. In this study the data collected is the data acquired by instrumental method in parallel study by Tanushree Sharma [38]. The data collected is the features extracted from the finger pulse profile of Photoplethysmograph. To acquire this data, three subjects were taken and the finger pulse waveform of three fingers (index, middle and ring) of both left and right hands were recorded (for the duration of 1 minute) using BIOPAC MP system and AcqKnowledge software well before lunch and immediately after lunch (as the Pitta level rises in day time after lunch). Each PPG record consists of more than 40 pulse waveforms of which a set of 25 pulses were selected and the pulse with more noise or undesired shape was discarded. From these 25 pulses, five features were extracted manually. These five features are:

a) TP1- time taken for the occurrence of first peak
b) TP2- time taken for the occurrence of second peak
c) TPT- total time duration of a pulse
d) A1- amplitude of first peak
e) A2- amplitude of second peak

These features are also shown in figure 2.6.

The hardware and software (BIOPAC MP System and AcqKnowledge) used to acquire the finger pulse waveform is briefly discussed below:

MP System: MP System is a complete and expandable data acquisition system. Functions of this system are like an on-screen chart recorder, oscilloscope, and X/Y plotter, allowing to record, view, save, and print data (Figure 5.1). It includes all the hardware and software required to turn any computer into a powerful data acquisition workstation specifically designed for life science applications. MP System is as powerful as larger and more expensive data acquisition systems, but has a familiar, easy to use graphical interface. This

System will reduce your equipment setup time and increase the quality of your results. By harnessing the power of your computer, the MP System gives you publication-quality results with minimum effort.

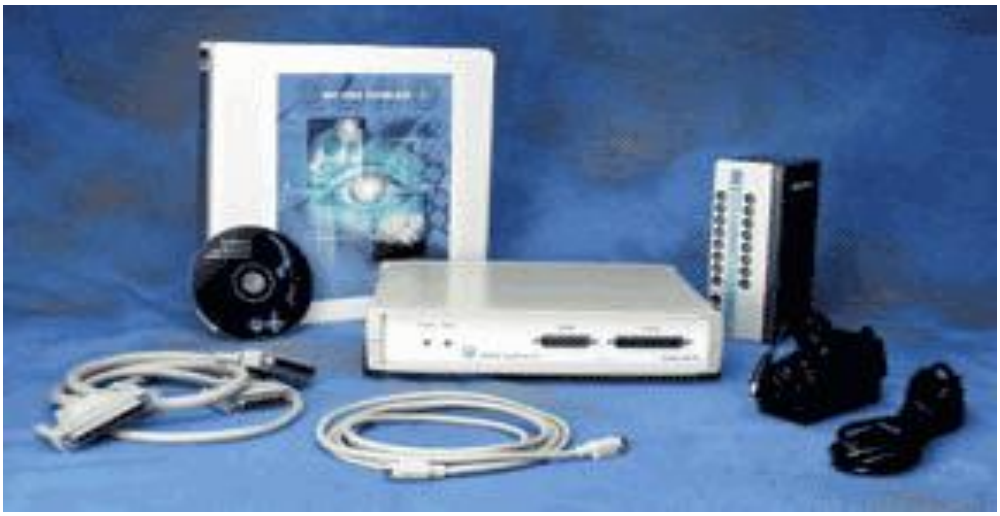

Fig 5.1 BIOPAC MP Data Acquisition Unit

The MP data acquisition unit (MP150 or MP100) is the heart of the MP System. The MP unit takes incoming signals and converts them into digital signals that can be processed with your computer. Data collection generally involves taking incoming signals (usually analog) and sending them to the computer, where they are (a) displayed on the screen and (b) stored in the computer‘s memory (or on the hard disk). These signals can then be stored for future examination, much as a word processor stores a document or a Statistics program saves a data file. Graphical and numerical representations of the data can also be produced for use with other programs [39]

AcqKnowledge: The MP System (MP150 or MP36R) software is called AcqKnowledge shown in figure 5.2. It performs two basic functions: acquisition and analysis. All of the acquisition parameters can be found under the MP150 menu. The other menus have commands related to analysis functions such as viewing, editing, and transforming data. Assuming everything is properly connected and there are no conflicts, Acq*Knowledge* opens graph window. A “window” is the term used for the area on the computer‘s screen where data is displayed and/or manipulated. The graph window on the screen is designed to provide you with a powerful yet easy-to-use interface for working with data [40].

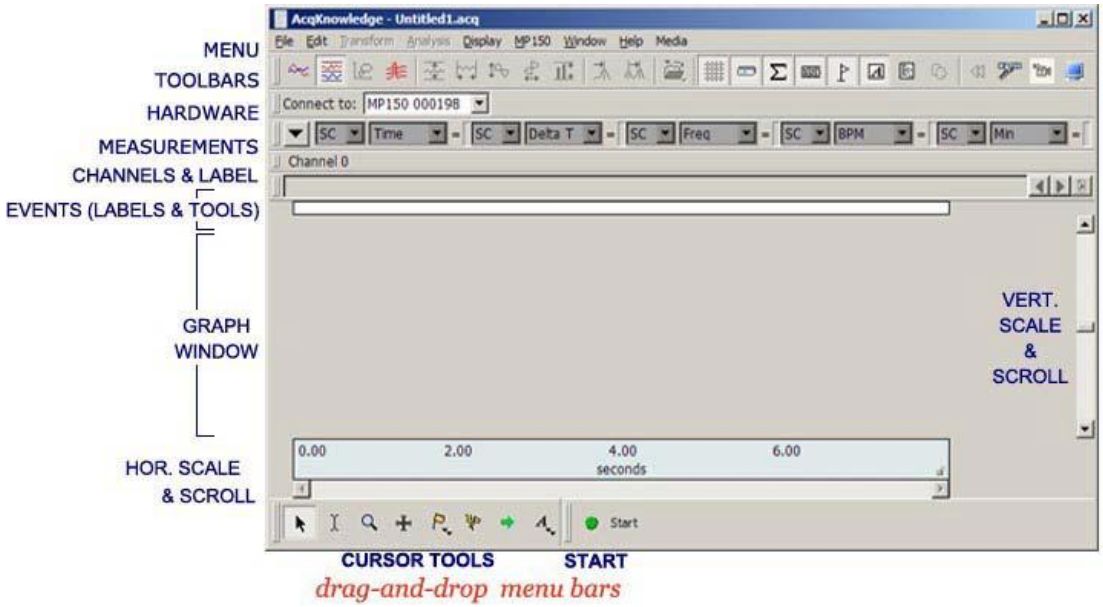

Fig. 5.2 AcqKnowledge (main window)

5.2 Feature Extraction

Once the data is collected, which consist of 25 pulses of each finger (index, middle and ring) of both left and right hands of three subjects of before and after lunch. Our next step is to extract the features from the data for the classification. The task of feature extraction is to get the most relevant information from the original data. The feature extraction is done by normalizing the five features of finger pulse waveform; the resulted values are TP1/TPT, TP2/TPT and A2/A1. To get the most relevant information, the mean and variance of these three normalized values are calculated. These calculated values are presented in a form of datasheets, which are shown below:

5.2.1 Subject 1 Left Hand (Index Finger)

Table 5.1 Features of left hand index finger (Before lunch)

	TP1/TPT	TP2/TPT	A2/A1
	0.215983	0.590844	0.284336
	0.215983	0.590844	0.261628
	0.214157	0.588819	0.270172
	0.190062	0.560079	0.299642
	0.218458	0.632139	0.284687
	0.206223	0.618524	0.296865
	0.217424	0.619622	0.272602
	0.216847	0.698917	0.164412
	0.215983	0.647791	0.266058
	0.217424	0.597803	0.291268
	0.206591	0.597803	0.266913
	0.210581	0.59997	0.267004
	0.225811	0.623698	0.264317
	0.200089	0.61061	0.245608
	0.227658	0.603892	0.28259
	0.206591	0.608789	0.26765
	0.224577	0.581638	0.247272
	0.216467	0.566922	0.303279
	0.219806	0.60435	0.261838
	0.234135	0.606391	0.285377
	0.230793	0.637458	2.588431
	0.215094	0.602113	0.287843
	0.24996	0.693093	0.267323
	0.222274	0.644517	0.300619
	0.217005	0.655211	0.283073
MEAN	0.217439	0.615273	0.364432
VAR	0.000136	0.001136	0.21542

Table 5.2 Features of left hand index finger (After lunch)

	TP1/TPT	TP2/TPT	A2/A1
	0.258917	0.61179	0.302151
	0.224952	0.63748	0.260114
	0.229913	0.574702	0.320706
	0.240994	0.6506	0.26546
	0.21681	0.590394	0.258832
	0.241368	0.620684	0.330939
	0.244165	0.627877	0.253006
	0.241407	0.586251	0.272837
	0.247028	0.623514	0.287208
	0.232577	0.592949	0.309144
	0.25	0.618984	0.261636
	0.244165	0.616289	0.272457
	0.252985	0.632139	0.273555
	0.238135	0.607119	0.259816
	0.224763	0.573028	0.255035
	0.238135	0.618984	0.225514
	0.273897	0.654746	0.248232
	0.252985	0.574702	0.336946
	0.212493	0.625022	0.225868
	0.234489	0.629636	0.276254
	0.258917	0.623514	0.293869
	0.222318	0.592965	0.264812
	0.246924	0.64183	0.261215
	0.243965	0.621983	0.267609
	0.238135	0.607119	0.289046
MEAN	0.240418	0.614172	0.27489
VAR	0.000196	0.000527	0.000811

5.2.2 Subject 1 Left Hand (Middle Finger)

Table 5.3 Features of left hand middle finger (Before lunch)

	TP1/TPT	TP2/TPT	A2/A1
	0.195778	0.536148	0.48582
	0.230888	0.560419	0.435884
	0.216568	0.525661	0.544455
	0.193788	0.5307	0.533398
	0.241794	0.587918	0.500542
	0.211002	0.544491	0.594486
	0.241457	0.563185	0.530657
	0.244292	0.592955	0.501202
	0.236465	0.602021	0.501959
	0.238725	0.613705	0.440347
	0.215936	0.579442	0.471698
	0.22356	0.599901	0.391757
	0.240547	0.645482	0.37164
	0.225909	0.548334	0.450328
	0.212741	0.563777	0.354437
	0.218846	0.531191	0.497372
	0.215051	0.569899	0.366523
	0.225909	0.569899	0.420369
	0.216568	0.567009	0.422264
	0.218846	0.563185	0.424741
	0.229218	0.583346	0.448186
	0.250038	0.597805	0.456024
	0.199911	0.557951	0.392927
	0.215051	0.559041	0.420146
	0.237097	0.556744	0.422058
MEAN	0.223839	0.570008	0.455169
VAR	0.000233	0.000807	0.003652

Table 5.4 Features of left hand middle finger (After lunch)

	TP1/TPT	TP2/TPT	A2/A1
	0.228259	0.554409	0.376491
	0.228259	0.587039	0.345384
	0.211139	0.555538	0.395718
	0.208838	0.593525	0.293047
	0.197964	0.552145	0.365094
	0.217332	0.565184	0.376418
	0.215015	0.56997	0.397707
	0.206556	0.565184	0.351137
	0.216496	0.567007	0.38705
	0.200029	0.53689	0.375876
	0.218764	0.572945	0.352532
	0.208291	0.562618	0.356685
	0.195629	0.543482	0.345946
	0.212715	0.563874	0.360795
	0.206102	0.546344	0.372774
	0.187491	0.552145	0.36767
	0.195912	0.536059	0.336954
	0.224533	0.561191	0.374329
	0.212684	0.585029	0.364391
	0.234791	0.551076	0.419958
	0.218732	0.562536	0.340899
	0.202139	0.56379	0.344727
	0.225623	0.580258	0.403155
	0.219666	0.582298	0.40286
	0.230746	0.593434	0.291875
MEAN	0.212948	0.564159	0.363979
VAR	0.000149	0.000263	0.000933

5.2.3 Subject 1 Left Hand (Ring Finger)

Table 5.5 Features of left hand ring finger (Before lunch)

	TP1/TPT	TP2/TPT	A2/A1
	0.214321	0.642798	0.510478
	0.173423	0.5	0.531872
	0.158427	0.494945	0.513003
	0.163416	0.480699	0.525334
	0.189488	0.526207	0.544752
	0.1851	0.524378	0.507933
	0.186231	0.490126	0.521946
	0.2	0.52	0.523773
	0.18993	0.49986	0.536733
	0.180922	0.485745	0.503921
	0.18993	0.50993	0.54059
	0.193834	0.50628	0.539495
	0.16993	0.5	0.522903
	0.180559	0.515117	0.519957
	0.1837	0.510277	0.512025
	0.158844	0.467381	0.530124
	0.163753	0.422404	0.58242
	0.184411	0.48547	0.549761
	0.198034	0.52472	0.518936
	0.186231	0.5	0.530391
	0.179203	0.5	0.506039
	0.188704	0.5	0.613024
	0.180922	0.485745	0.526467
	0.192308	0.5	0.600832
	0.201856	0.480769	0.558687
MEAN	0.183739	0.502914	0.534856
VAR	0.000186	0.001313	0.000787

Table 5.6 Features of left hand ring finger (After lunch)

	TP1/TPT	TP2/TPT	A2/A1
	0.202296	0.607054	0.367085
	0.202296	0.607054	0.358312
	0.197595	0.592948	0.323348
	0.197595	0.592948	0.313312
	0.224682	0.584236	0.312752
	0.211738	0.600033	0.272061
	0.227251	0.590916	0.299223
	0.241285	0.620723	0.239074
	0.224717	0.595477	0.322109
	0.227251	0.590916	0.331579
	0.227251	0.590916	0.302539
	0.223574	0.599868	0.347668
	0.230687	0.593457	0.345919
	0.238526	0.613626	0.315281
	0.208845	0.560504	0.312421
	0.224682	0.550636	0.333866
	0.199901	0.564688	0.297936
	0.208845	0.538544	0.280202
	0.226252	0.618865	0.273691
	0.246958	0.635317	0.273909
	0.235284	0.599967	0.262496
	0.227143	0.626368	0.293911
	0.214274	0.618865	0.32249
	0.234645	0.641822	0.242273
	0.222222	0.629572	0.260856
MEAN	0.221032	0.598613	0.304173
VAR	0.000199	0.000651	0.001201

5.2.4 Subject 1 Right hand (index finger)

Table 5.7 Features of Right hand index finger (Before lunch)

	TP1/TPT	TP2/TPT	A2/A1
	0.2045	0.5227	0.560044
	0.2472	0.5842	0.458656
	0.2174	0.5217	0.469905
	0.2197	0.5824	0.435199
	0.2028	0.4904	0.501746
	0.23	0.5299	0.522801
	0.2339	0.5532	0.488733
	0.2246	0.5618	0.490432
	0.2223	0.5555	0.458333
	0.2199	0.5199	0.565643
	0.2392	0.5761	0.533662
	0.2325	0.5698	0.488732
	0.187576	0.525061	0.44848
	0.176529	0.517581	0.463155
	0.185241	0.543273	0.462349
	0.178623	0.5	0.417454
	0.199971	0.49414	0.453055
	0.190483	0.51186	0.440778
	0.215824	0.511323	0.522271
	0.195391	0.494274	0.526916
	0.195391	0.494274	0.522626
	0.18825	0.517581	0.490374
	0.18607	0.5	0.488202
	0.192798	0.530083	0.44989
	0.190483	0.5	0.551276
MEAN	0.207065	0.528282	0.488428
VAR	0.00042	0.000877	0.001674

Table 5.8 Features of Right hand index finger (After lunch)

	TP1/TPT	TP2/TPT	A2/A1
	0.220825	0.546583	0.496221
	0.211029	0.522335	0.466921
	0.189002	0.500154	0.450138
	0.181947	0.5115	0.454611
	0.20922	0.534816	0.444808
	0.199967	0.576578	0.418424
	0.22611	0.595148	0.416872
	0.204777	0.590446	0.426214
	0.228829	0.590446	0.415089
	0.215125	0.595017	0.401314
	0.200139	0.612545	0.404014
	0.202667	0.60765	0.424734
	0.21069	0.631704	0.363874
	0.202847	0.635138	0.380154
	0.20526	0.628221	0.425362
	0.213454	0.639993	0.391453
	0.21069	0.61857	0.342113
	0.22676	0.639993	0.353547
	0.220882	0.623402	0.352958
	0.233663	0.623402	0.342645
	0.213309	0.640111	0.35115
	0.215339	0.620218	0.36539
	0.240435	0.633029	0.370006
	0.239882	0.639993	0.329486
	0.223823	0.61857	0.325429
MEAN	0.213867	0.599022	0.396517
VAR	0.000208	0.00186	0.002148

5.2.5 Subject 1 Right hand (middle finger)

Table 5.9 Features of Right hand middle finger (Before lunch)

	TP1/TPT	TP2/TPT	A2/A1
	0.26664	0.552411	0.506832
	0.263142	0.6	0.430276
	0.258779	0.623564	0.424272
	0.232352	0.565591	0.387336
	0.23593	0.617965	0.32877
	0.255799	0.651095	0.312463
	0.24412	0.639416	0.317454
	0.211718	0.599869	0.274124
	0.231711	0.646308	0.268888
	0.218409	0.643682	0.288783
	0.258026	0.644914	0.343076
	0.238	0.64275	0.30548
	0.215916	0.625079	0.278621
	0.204874	0.650588	0.263338
	0.21348	0.618025	0.302562
	0.212791	0.585102	0.281661
	0.197908	0.55202	0.27066
	0.183772	0.520427	0.281589
	0.212061	0.535297	0.310148
	0.217926	0.57423	0.254089
	0.224626	0.591886	0.291573
	0.232352	0.575736	0.300669
	0.17117	0.51351	0.593971
	0.268827	0.537654	0.672224
	0.255089	0.520427	0.573338
MEAN	0.229017	0.593102	0.354488
VAR	0.000672	0.002116	0.013285

Table 5.10 Features of Right hand middle finger (After lunch)

	TP1/TPT	TP2/TPT	A2/A1
	0.227822	0.620297	0.410901
	0.212545	0.61258	0.329973
	0.229052	0.602532	0.329719
	0.229052	0.626687	0.296096
	0.216892	0.614526	0.311932
	0.218008	0.653846	0.333011
	0.214286	0.595293	0.315423
	0.212545	0.61258	0.304324
	0.229052	0.614526	0.298122
	0.222222	0.62963	0.300605
	0.222222	0.605052	0.282776
	0.212545	0.625022	0.291846
	0.209933	0.617341	0.286686
	0.204898	0.614526	0.340289
	0.215223	0.620297	0.26963
	0.224987	0.637463	0.288977
	0.215223	0.632896	0.2929
	0.237602	0.650078	0.32474
	0.199931	0.625022	0.342087
	0.227822	0.632896	0.388243
	0.199931	0.637463	0.299516
	0.20245	0.632896	0.274307
	0.205069	0.641085	0.28915
	0.204898	0.602532	0.280402
	0.205069	0.628323	0.265669
MEAN	0.215971	0.623416	0.309893
VAR	0.000112	0.000218	0.001205

5.2.6 Subject 1 Right hand (ring finger)

Table 5.11 Features of Right hand ring finger (Before lunch)

	TP1/TPT	TP2/TPT	A2/A1
	0.360178	0.520903	0.409531
	0.363545	0.550062	0.389927
	0.36	0.531836	0.436131
	0.354353	0.500072	0.458012
	0.339623	0.552003	0.402428
	0.358491	0.557813	0.378155
	0.346939	0.532609	0.504638
	0.354064	0.648715	0.32159
	0.352941	0.614458	0.34889
	0.34	0.588139	0.318092
	0.327861	0.616186	0.293224
	0.377612	0.581009	0.406548
	0.390432	0.556621	0.428519
	0.349861	0.576828	0.462385
	0.344846	0.585837	0.394593
	0.344846	0.585837	0.428201
	0.344846	0.585837	0.339677
	0.344846	0.585837	0.426788
	0.344846	0.574219	0.425856
	0.344846	0.563052	0.425319
	0.344846	0.597935	0.38744
	0.357182	0.583307	0.378844
	0.327263	0.591384	0.410638
	0.33324	0.60682	0.377221
MEAN	0.350313	0.574472	0.398027
VAR	0.000202	0.001103	0.002415

Table 5.12 Features of Right hand ring finger (After lunch)

	TP1/TPT	TP2/TPT	A2/A1
	0.212504	0.625044	0.267474
	0.222203	0.641977	0.244291
	0.24434	0.608119	0.240952
	0.226125	0.64275	0.2497
	0.20688	0.62064	0.265523
	0.21425	0.631042	0.273252
	0.204842	0.650584	0.258032
	0.202375	0.64275	0.270616
	0.22905	0.674793	0.270663
	0.19768	0.60464	0.301731
	0.209887	0.641977	0.257416
	0.224974	0.662452	0.263461
	0.212504	0.637513	0.254531
	0.222203	0.678925	0.27137
	0.222203	0.641977	0.268485
	0.200035	0.637513	0.307021
	0.224974	0.649982	0.259843
	0.222203	0.641977	0.259965
	0.228881	0.626545	0.298329
	0.215188	0.65819	0.240801
	0.237619	0.649982	0.232719
	0.213575	0.629203	0.326644
	0.231837	0.646333	0.301521
	0.207847	0.675365	0.27322
	0.204842	0.638564	0.308591
MEAN	0.217561	0.642353	0.270646
VAR	0.000144	0.000346	0.000578

The mean and variance of other two subjects are calculated in the same way. It can be seen that there is no visible change in the mean values of the parameters before and after lunch;

however the variance tends to decrease in some of the parameters, in most of the subjects. To confirm this change, a statistical check needs to be performed. For the kind of data presented, t-test is chosen for this purpose. .

5.3 Feature Selection

A good classifier is constructed only if the appropriate set of features is selected because irrelevant features simply add noise to the data and affect model accuracy. The best feature has been selected on the basis of the change in its value after taking the meal in daytime. And it is the matter of great importance that the change in value of extracted feature must be statistical significant. Hence, for the selection of most appropriate feature t-test is applied on the variance (before and after lunch) of each normalized value of each finger of both hands of all three subjects. The purpose of t-test is to find whether the change in value of feature after the lunch is just by chance or some factor is responsible for the change. Thus, the t-test will return the most appropriate feature which shows the significant change in the level of Pitta after the meal is consumed in daytime. The process described above has been implemented and the results have been discussed in the proceeding chapter of this thesis.

Chapter-6

RESULTS AND DISCUSSION

As it is observed in the previous chapter, features extracted from the pulse profile were analysed for their mean value as well as for their variance in the 25 subsequent valid waves. There is no visible difference in the values of mean therefore all the values of mean are not considered in further study. However, the values of variance that are arrived at had prominent difference, before and after the lunch. In our pilot study, three subjects were considered. The readings were taken from three fingers (index, middle, and ring) of both left hand right hands of each subject. Each finger has three features namely A2/A1, TP1/TPT, and TP2/TPT. And these three features were recorded before and after the lunch. Therefore we have total of 108 values of features are considered.

To apply the t-test the 108 values of variance are considered in a set of three features and that too in a set of before and after lunch, therefore a total of $\frac{108}{(2\times3)} = 18$ sets are considered for calculating the t-test. The t-test results of these 18 sets are shown below from table 6.1 to 6.6. In these tables the value of t-test actually shows the value of *p,* if there is any significant change in the value of variance after lunch the *p* value must come to be less than 0.05.

6.1 t-test results

Table 6.1 t-test values of left hand index finger

	feature	Subject 1	Subject 2	Subject 3	t-test	consistency
Left hand index finger						
Before lunch	A2/A1	0.21542	0.000757	0.000757	0.423935	no
After lunch	A2/A1	0.000811	0.001006	0.000985		
Before lunch	TP1/TPT	0.000136	0.000173	0.000173	0.170918	yes
After lunch	TP1/TPT	0.000196	0.000204	0.000179		
Before lunch	TP2/TPT	0.001136	0.000318	0.000318	0.740939	no
After lunch	TP2/TPT	0.000527	0.001296	0.000471		

Table 6.2 t-test values of left hand middle finger

	feature	Subject 1	Subject 2	Subject 3	t-test	consistency
Left hand middle finger						
Before lunch	A2/A1	0.003652	0.001632	0.001632	0.11755	yes
After lunch	A2/A1	0.000933	0.000421	0.000863		
Before lunch	TP1/TPT	0.000233	0.00022	0.00022	0.887153	no
After lunch	TP1/TPT	0.000149	8.97E-05	0.000389		
Before lunch	TP2/TPT	0.000807	0.000694	0.000694	0.006359	no
After lunch	TP2/TPT	0.000263	0.000795	0.000622		

Table 6.3 t-test values of left hand ring finger

	feature	Subject 1	Subject 2	Subject 3	t-test	consistency
Left hand ring finger						
Before lunch	A2/A1	0.000787	0.000312	0.007782	0.522816	no
After lunch	A2/A1	0.001201	0.000598	0.0036		
Before lunch	TP1/TPT	0.000186	0.00023	0.000224	0.412263	no
After lunch	TP1/TPT	0.000199	0.000213	9.13E-05		
Before lunch	TP2/TPT	0.001313	0.000425	0.001483	0.13851	yes
After lunch	TP2/TPT	0.000651	0.000265	0.000416		

Table 6.4 t-test values of right hand index finger

	feature	Subject 1	Subject 2	Subject 3	t-test	consistency
Right hand index finger						
Before lunch	A2/A1	0.001674	0.000305	0.001764	0.929037	no
After lunch	A2/A1	0.002148	0.000464	0.001021		
Before lunch	TP1/TPT	0.00042	7.82E-05	0.00015	0.412596	no
After lunch	TP1/TPT	0.000208	0.000143	4.9E-05		
Before lunch	TP2/TPT	0.000877	0.000567	0.000451	0.370684	no
After lunch	TP2/TPT	0.00186	0.000832	0.000322		

Table 6.5 t-test values of right hand middle finger

	feature	Subject 1	Subject 2	Subject 3	t-test	consistency
Right hand middle finger						
Before lunch	A2/A1	0.013285	0.000977	0.007723	0.196125	yes
After lunch	A2/A1	0.001205	0.000633	0.000651		
Before lunch	TP1/TPT	0.000672	0.000124	0.000185	0.607236	no
After lunch	TP1/TPT	0.000112	0.000187	0.000291		
Before lunch	TP2/TPT	0.002116	0.000557	0.000301	0.616157	no
After lunch	TP2/TPT	0.000218	0.000676	0.000777		

Table 6.6 t-test values of right hand ring finger

	feature	Subject 1	Subject 2	Subject 3	t-test	consistency
Right hand ring finger						
Before lunch	A2/A1	0.002415	0.001779	0.000825	0.12998	yes
After lunch	A2/A1	0.000578	0.001066	0.000303		
Before lunch	TP1/TPT	0.000202	0.00027	0.000105	0.369956	no
After lunch	TP1/TPT	0.000144	0.000286	7.17E-05		
Before lunch	TP2/TPT	0.001103	0.000924	0.00072	0.023373	yes
After lunch	TP2/TPT	0.000346	0.00047	0.000186		

From The t-test values shown above (Table 6.1 to Table 6.6) only those values were considered for further analysis which showed the consistent increase or decrease in the value of variance after the lunch in both hands. These values are A2/A1 feature of middle finger of both hands and the TP2/TPT feature of ring finger of both hands. These two features showed the consistent decrease in value of variance after the lunch. The values used for further analysis are shown in table 6.7.

Table 6.7 t-test values of selected features

	Feature	Subject_1	Subject_2	Subject_3	t-test
Left hand middle finger					
Before lunch	A2/A1	0.0036518	0.0016317	0.0016317	0.1175502
After lunch	A2/A1	0.0009334	0.0004211	0.0008632	
Right hand middle finger					
Before lunch	A2/A1	0.0132852	0.0009770	0.0077226	0.1961249
After lunch	A2/A1	0.0012046	0.0006329	0.0006512	
Left hand ring finger					
Before lunch	TP2/TPT	0.0013129	0.0004252	0.0014831	0.1385104
After lunch	TP2/TPT	0.0006509	0.0002654	0.0004161	
Right hand ring finger					
Before lunch	TP2/TPT	0.0011028	0.0009244	0.0007196	0.0233727
After lunch	TP2/TPT	0.0003463	0.0004704	0.0001855	

From the Table 6.7 it is observed that from all four t-test results only one t-test result shows $p< 0.05$ and i.e. 0.0233727. This value shows that only TP2/TPT feature of ring finger of right hand is proved to be statistically significant. Hence, the change occurred in the value of TP2/TPT feature after lunch is not by chance and there is some factor responsible for it. The factor responsible may be the rise of the Pitta level in subject's body immediately after lunch. However, in another approach if the sets of left hand and right hand of before and after lunch are combined together the probability value of t-test further decreases which is shown in Table 6.8.

Table 6.8 t-test results of variance of selected features from combination of left and right hand

		Left hand			Right hand			
Feature		**Subjec1**	**Subjec2**	**Subjec3**	**Subject1**	**Subject2**	**Subject3**	**t-test**
A2/A1 feature of middle finger								
Before lunch	A2/A1	0.00365	0.00163	0.00163	0.01328	0.00097	0.00772	0.08688
After lunch	A2/A1	0.00093	0.00042	0.00086	0.00120	0.00063	0.00065	
TP2/TPT feature of ring finger								
Before lunch	TP2/TPT	0.00131	0.00042	0.00148	0.00110	0.00092	0.00071	0.00463
After lunch	TP2/TPT	0.00065	0.00026	0.00041	0.00034	0.00047	0.00018	

It is observed from the above study that variance in TP2/TPT of ring finger of the subjects is a possible candidate for detecting higher Pitta levels.

Chapter-7

CONCLUSION AND FUTURE SCOPE

7.1 Conclusion

On account of increased Pitta level after the lunch some of the selected features A2/A1 of middle finger and TP2/TPT of ring finger of both left and right hands of all the three subjects showed a consistent decrease in its variance. However, the statistical analysis (t-test) of this data showed that this change is significant only in TP2/TPT feature of ring finger of right hand. Therefore it is concluded that TP2/TPT feature of ring finger can be used as one of the best feature for the classification of high Pitta level in body after taking the lunch.

7.2 Future Scope

This study is limited to only three subjects. This being a pilot study has provided some insight into the changes that occur in the selected features of finger pulse profile. However, this study needs to be extended to a larger no. of subjects say about 36 or so for the data acquired under different conditions. More parameters need to be analysed like that extracted from second time derivative of the signal and from that extracted in frequency domain.

REFERENCES

[1] "Scientific classification," *http://en.wikipedia.org/wiki/Biological_classification.html.*

[2] D. Michie, D.J. Spiegelhalter and C.C. Taylor, *Machine Learning, Neural and Statistical Classification.* Ellis Horwood Series in Artificial Intelligence, February 17, 1994.

[3] Kirk Allen, *Explaining Chronbach's alpha,* Department of Industrial Engineering, The University of Oklahoma.

[4] "Feature extraction and selection methods," *http://isa.umh.es/asignaturas/cscs/PR/3-Feature extraction.pdf.*

[5] L.Yu and H. Liu, Efficient Feature Selection via Analysis of Relevance and Redundancy, *JMLR*, 5(Oct):1205-1224, 2004.

[6] Markovitch S. & Rosenstein D. (2002), Feature Generation Using General Construction Functions, *Machine Learning* 49: 59-98.

[7] Michael S. Rosenberg, "A Generalized Formula for Converting Chi-Square Tests to Effect Sizes for Meta-Analysis," *PLoS ONE 5(4): e10059.doi:10.1371/journal.pone. 0010059,* 7th April,2010.

[8] James P. Key, "Research Design in Occupational Education," http://www. okstate.edu /ag/agedcm4h/academic/aged5980a/5980/newpage26.html, 1997.

[9] Charles Elkan, *Evaluating Classifiers,* University of California at San Diego, January 20,2012.

[10] Robert R. Bailey, *Classifier Training and Evaluation*, National Taiwan Normal University, *www.csie.ntnu.edu.tw/~bbailey/ClassifTrainingEval.pdf.*

[11] V. Cherkassky and F. Mulier, Learning from Data: Concepts, Theory, and Methods. *Wiley Interscience*, 1998.

[12] S. Cost and S. Salzberg, "A Weighted Nearest Neighbour Algorithm for Learning with Symbolic Features," *Machine Learning*, 1993.

[13] P. Langley, W. Iba, and K. Thompson, "An analysis of Bayesian classifiers," *In Proc. Of the 10th National Conf. on Artificial Intelligence,* 1992.

[14] Xu Dezhi and Gamage Upeksha Ganegoda, "Rule Based Classification to Detect Malnutrition in Children." *International Journal on Computer Science and Engineering (IJCSE)*, January 2011.

[15] "Nearest neighbour search", *http:// en.wikipedia.org/wiki/Nearest_neighbor_search.html*

[16] "Classifier Showdown", *http://blog.peltarion.com/2006/07/10/classifier-showdown.html*

[17] "Artificial Neural Networks", HCMC University of Technology, September, 2008.

[18] Mingyue Tan, *Support Vector Machine & its Applications*, The University of British Columbia, November 26, 2004.

[19] "Ayurveda: A Brief History", *http://www.allayurveda.com*

[20] S.Sinha, *http://banglore.metromela.com/ Expert+guide+on+Nadi+Pariksha /article /39 98.*

[21] G. Chan, P.M. Middleton, N.H. Lovell and B.G. Celler, "Extraction of photoplethysmographic waveform variability by lowpass filtering," *Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, Shanghai, China, September 1-4,2005.

[22] John Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiological Measurement, IOP Publishing*, Vol. 23, pp. 1-39, January,2007.

[23] R. Gonzalez, A. Manzo , J. Delgado, J.M. Padilla, B. Trenor, and J. Saiz, "A computer based photoplethysmographic vascular analyzer through derivates," *Comput. Cardiol.* 35: 177–180, 2008.

[24] D. Korpas, J. Halek and L. Dolezal, "Parameters Describing the Pulse Wave," *Physiol. Res.* 58, pp. 473-479, 2009.

[25] Richard Mankiewicz, *The Story of Mathematics*. Princeton University Press, p.158, 2001.

[26] Eivind Hoffman, "Standard Statistical Classifications: Basic Principles," *STATISTICAL COMMISSION,* Thirtieth session, New York, 10 February, 1999.

[27] W.Duch, N. Janowski, K. Grabczewski and R. Adamczak, "Optimization and Interpretation of Rule – Based Classifiers," *Intelligent Information Systems IX,* 2007.

[28] D. LU and Q. Weng, "A Survey of Image Classification Methods and Techniques for Improving Classification Performance," *International Journal of Remote Sensing,* Volume 28, No.5, pp. 823-870, March,2007.

[29] G. Tsoumakas and I. Katakis, "Multi-label Classification: An Overview," *International Journal of Data Warehousing and Mining,* Idea Group Publishing, pp 1-13, 2007.

[30] S. B. Kotsiantis, "Supervised Machine Learning: A review of Classification Techniques," *Informatica* 31, pp. 249-268, July 16, 2007.

[31] Mohsen Tavalok and Reg Dennick, "Making Sense of Chronbach's alpha," *International Journal of Medical Education,*2011.

[32] L Mishra, BB Singh, S Dagenais, "Healthcare and Disease Management in Ayurveda," *Alternative Therapies in Health and Medicine,* Volume 7, issue 2, pp.44-50, 2001

[33] Sandrine *et al.*, "Contour analysis of the photoplethysmographic pulse measured at the finger," *Journal of Hypertension*, vol. 24, no. 8, pp1449–1456, 2006.

[34] Zbignevs Marcinkevics, Signe Kusnere, Juris Imants Aivars, Uldis Rubins and Aram Hussain Zehtabi, "The shape and dimensions of photoplethysmographic pulse waves: A Measurement Repeatability Study", *Acta Universitatis Latviensis,* Biology, vol. 753, pp. 99–106.

[35] G.M. Kadhar Nawaz *et al*, "E- Health Analysis Element for Supporting Therapeutic through Ancient Indian Medical Science," *International Journal of Computer Applications,* Volume 1- No.17, 2010.

[36] Mandeep Singh, and Spiti Gupta, "Correlation Studies of Finger Pulse Profiles for Detecting Ayurvedic Doshas," *International Journal of Computer Science and Communication*, vol. 2, no. 2, pp. 373-375, July-December 2011.

[37] John Joseph Immel, "The Three Doshas," *http://www.joyfulbelly.com/Ayurveda/article /dosha/Pitta.html, 2011.*

[38] Tanushree Sharma, "Feature Extraction from Finger Pulse Profile for High Pitta Detection," M.E. Thesis, Thapar University, EIED, Patiala, 2012.

[39] "MP System Hardware Guide", *http://www.biopac.com/LabManuals.asp.*

[40] "AcqKnowledge Software Guide", *http://www.biopac.com/LabManuals.asp.*

PUBLICATIONS

Papers Accepted

1. Mandeep Singh and Bharti Chauhan, "Classification: A holistic view," *International Journal of Computer Science and Communication*, vol. 3, no. 2, 2012 (in press).

2. Mandeep Singh and Bharti Chauhan, "High Pitta Detection using Finger Photoplethysmograph based Features: A Feasibility Study," *International Journal of Computer Science and Communication*, vol. 3, no.2, 2012 (in press).