CHAPTER – I
CHAPTER - I

INTRODUCTION TO PROBLEM

1.1. INTRODUCTION

The mining of large databases and making classifications, associations and predictions on many scientific aspects can highlight many latent facts from the past data or existing information. Mathematical methods of Artificial Neural Networks, Fuzzy logic and data mining rules when applied on these databases can bring out new knowledge. This is an age of marvelous complexity and variety, a world where events are never exactly the same, they are also not completely different, only approximation techniques can be used to analyze the databases.

The data mining and neuro computing algorithm show a thread of continuity, similarity and predictability that allows us to generalize from the past experience to future events. The mathematical algorithms used here deal with the correlation between the various factors causing diabetes mellitus and the extent to which each of the factor affects the individual in the data sample. The uncertainty, risk factors due to various alleles of the genes can be modelled with fuzzy logic and probabilistic coefficient to draw the relationship between each factor and their total effect in causing the disease.

Neural networks similar to human brain can show intrinsic generalization in the categorization of objects. Fuzzy or multivalued set theory
developing the basic insight the categories are not absolutely clear cut: In most real world problems a particular example of a category or a class can “belong” to “lesser” or “greater” degrees to that category. Computer based analysis [1] of medical databases can give pertinent clinical details at any point of emergency. The enhanced and redefined precision of diagnostic details improved the quality of medical care in the sacrosanct world of medicine. Even genetic pre-disposition of hereditary diseases like diabetes mellitus are possible with computer aided predictions in human genes.

Neural networks [2] and fuzzy systems estimate input-output functions. Both are trainable dynamical systems. Sample data shapes and “program” their time evolution. They are model-free estimators. They “learn from experience” with numerical and sometimes linguistic sample data. Artificial neural networks consist of numerous, simple processing of units or “neurons” that we can globally program for computation [3]. Programming or training neural networks can store, recognize and associatively retrieve patterns or database entries.

In chapter one of the thesis, a brief introduction of the theme is given. In chapter two a survey of what other researchers have contemplated has been summarized followed by their limitations too along with the various modules and block diagrams. In chapter three pathogenesis of the disease with their genetic and phenotypic factors in various ethnic group are discussed in different types of Diabetes Mellitus. In chapter four fuzzy logic application
model has been described and its Annexure gives the Table with data enlightening the degree of various risky genes inherited. The model has been described with results and discussions for predicting the genetic risk behind the disease. In chapter five the mathematical model has been projected on the problem and reasonable interpretations have been given followed by the results and discussions. In chapter six Neural Network model has been discussed in parallel with the mathematical modelling. Supervised learning technique is adopted since the target output is known (diabetic or non diabetic) and parameters are iterated towards the knowledge about acquiring of diabetes and knowing the genetic risk of the same. In chapter seven Data mining model has been arrived at using Bayesian probability, logistic regression and ID3 algorithm to analyze the diabetes mellitus in pima Indian sample. In chapter eight the various exhibits of the findings about Diabetes Mellitus using computer techniques have been given. In chapter nine the entire discussion is concluded giving the gist of the full process of various findings. It is followed by Bibliography.

1.2. IDENTIFICATION OF THE PROBLEM

The real Genome world is large and complex. Complexity arises from uncertainty in the form of abstractness of the patterns. Human beings address problems featuring complexity and ambiguity subconsciously. The main aim is to develop a model based on computers to perform the task of artificial
thinking and analysis of the complex human Genome in relation to hereditary diseases like Diabetes Mellitus. The main objective is to identify the genes related to Type I Diabetes Mellitus in various ethnic groups and causes of Type II diabetes in Pima Indian Population of Women. Genes are generally represented in terms of nucleotides (DNA & RNA) in the sequence of Adenine, Guanine, Thymine & Cytosine. Any change brought about in the sequence of the base pairs leads to mutations. The goal is centered towards finding pattern among gene sequences, which have undergone mutations leading to Diseases. Mutations may be due to Radiation, chemical agents and Replication errors even temperature leading to Fuzzy categories of the results.

The work also aims at a clinical study of actual patients acquiring diabetes mellitus both by hereditary factors as well as by phenotypic characteristics. Data have been collected from about 768 Indian Origin females who were tested for the presence of diabetes mellitus of which 268 were tested to be positive. The significant factors like the frequency of pregnancy, plasma glucose concentration in saliva, blood pressure variations due to change the environmental and physical factors, Forceps skin fold thickness, the insulin concentration in serum the obesity of patients due to lack of exercise, fat consumption, hereditary factors, age are playing vital role in the acquiring of diabetes mellitus in pima Indian population. A statistical model has been developed and significant conclusions about the dominance of each factor in
the cause of the disorder and statistical coefficients are calculated to substantiate the analysis.

**Figure [1.1] Types of Diabetes Mellitus [4]**

In Diabetes Mellitus IDDM is *multigenic*, multifactor disease fig[1.1] and it is hence difficult to produce a model mimicking human IDDM to understand the role played by these individual alleles and their associated significances. The major Histocompatibility computer (MHC) genes have been implicated in susceptibility and protection against IDDM. The recent advances in molecular biology have helped to find out the specific genes associated with IDDM. HLA-DR and HLA-DQ alleles have drawn attention to link their role in IDDM pathogenesis [5]. These investigations may prove beneficiary from the field of computer science.