CHAPTER – II
CHAPTER – II
SURVEY OF LITERATURE AND MODULES

2.1. LITERATURE SURVEY

The entire study was logically carried out with the first review on diabetes which is going to be a threat amidst all killer diseases that our mankind would face the ensuing future. Home page of the American diabetes association http://www.diabetes.org [6] gave a complete idea on how the disorder develops with the basic symptoms of excessive false hunger, polyurination, frequent sweating, abnormal thirst and sudden loss of weight in the individuals. The study revealed that the disorder caused in the beta cells of pancreas leads to this disorder. Next the types of diabetes and their role in causing the imbalance are Glucose metabolism of the body is studied. www.diabetes.ca/about diabetes/index.html [7] gave the mechanisms by which type and type 2 diabetes was caused. The methoddisthealth.org [8] was also instrumental in giving statistics about diabetes. The Home page for the National Institute of Diabetes and Digestive and Kidney diseases [9] gave a complete review on various parameters causing the disorder. The clinical study of the diabetes mellitus was carried out and analysis of the results by mathematical modelling led to various new findings about the disease which can be taken as a vital preventive or precautionary step towards the control of the same. An awareness about the disease could be achieved by the same. The
paper on “MHC CLASS II ALLELES AND INSULIN FROM NORTH INDIA”. HLA DR3, WHEN IN COMBINATION WITH DR4 IS A MARKER FOR THE EARLY DEVELOPMENT OF ISLET AUTOIMMUNITY by P.G. Colman, C. Steele, P. Wraight, J. Couper, S. Beresford, T. Powell, B.D. Tait, C. Harrison Walter & Eliza Hall Institute and Department of Diabetes and Endocrinology, Royal Melbourne Hospital, Victorian Transplantation and Immuno-genetic service, Melbourne, women’s and children hospital Adelaide [10], gave the complete risk levels of DR3 and DR4 genes in causing the disease to a greater extent.

The paper on ‘DNA IMMUNIZATION TO TREAT AUTOIMMUNE DIABETES” Matihias G. Von, Herrath, Advsem Bot, J. Lmdsay Whitton and Bryon Coon Depts. of Neuropharmacology and Immunology, Scripps Research Institute; La Jolla, [11] gave an idea on how autoimmunity could cause diabetes disorder without any previous indications, which could suddenly make a patient chronic. The paper on AUTO ANTIBODIES AND HLA SUSCEPTIBILITY MARKERS IN CANADA FIRST DEGREE RELATIVES OF PATIENTS WITH TYPE I DIABETES”. M.L. Beheme, J.L. Mahon and J. Purpre for the Canadian participate in ENDIT London Health science center. London, ONCANADE [12] gave a complete view of how the HLA susceptibility markers play a significant role in this pancreatic disorder. The paper HLA-DR4 BINDING REQUIREMENTS OF NATURALLY PROCESSED IA-2 PEPTIDES” by S. Arif, H. Coright and M. Peakman of
Department of Immunology, GKT School of medicine Kmg’s college London, UK [13] gave a complete literature study on how HLA-DR4 genes and autoimmunity affect the pancreatic functioning and the instability is caused in the Glucose metabolism of the affected individuals. The website http://diabetes-in-america.s-3.com [14] gave lot of information about Diabetes, its symptoms, consequences and the methods of bringing control by change in life style, food habits etc. The paper on Genetics of diabetes by the home page of WEHI gave an impression that IDDM is a complex genetic disease [15]. The contribution to the risk of diabetes is equally shared between environmental and genetic factors. The paper on “Genes can cause type 3 diabetes by Amy Adams. MS reviewed by Jeremy Walston, MD reviewed by Jeremy Walston, MD, updated sep15, 2000, GENETIC HEALTH [16] gave significant methods of identifying type 2 diabetes genes. The role of Beta 2 Adrenergic Receptor gene and the way of mutations causing the sudden proneness to this disease. The TRP64ARG mutation leading to type 2 diabetic with characteristics symptoms like more obese, slower metabolism, having a hard time losing weight, develops diabetes at an earlier age. The Huge review published in Epidemiologic Reviews 2000, 22(2): 218:27 [17] stressed the importance of HLA-DQ and type 1 DIABETES. It also gave a complete picture on the significant role of genes in the human leukocyte Antigen region on the short arm of chromosome no. 6. The DQ locus with tow tightly linked genes, DQA1 and DQB1 that encode alpha and beta glycoproteins with their gene
variants like DQA1 that encode alpha and beta glycoproteins allele DQB1 alleles, DQA1-DQB1 haplotypes the interactions of HLA-DQ AND Dr genes which are powerful in causing the disorder in various ethnic group like Caucasians, Asians, Japanese and other origins. The study of risk levels of these alleles were developed using fuzzy membership graph and extent of risk of both protective and risky genes are demonstrated with graphs.

The paper on HLA system and IDDM by Tiven Marwab and Pierre Y. Benhamou, Grenoble, Julliet 1996 [18] gave a complete picture of HLA genes the class I, Class II and Class III genes. The paper led to a finding that the distribution of certain Human Leucocyte antigens (HLA) has been found to be significantly different in diabetic patients compared to the healthy population. This observation and the tendency of IDDM to run in families indicate that the development of this auto immune response is genetically determined. The alleles found at the different loci of a haplotype, which are not distributed just by chance, but according to certain preferred association called “Linkage disequilibrium” forms a strong association with this disease. This was also observed from this paper. The paper also revealed a fact that frequency of Class II antigens DR3 and DR4 were much more increased among IDDM patients compared to healthy control population, than the Class I antigens. The negative association caused by genes like DR2 were also found from this paper. The paper on “PATHOGENESIS OF TYPE I A DIABETES, by George Elisenbarth, Chap 5 of March 2000 edition (20) [14] revealed that
there are series of stages in the disease beginning with genetic susceptibility and ending (from an immunologic stand point) with complete islet beta cell destruction.

The paper [19] also gave an idea on the Hierarchy of HLA terminology with specific alleles, groups of alleles, of different genes on the same chromosome, and genotypic variations. The Hierarchy of diabetes risk with examples of haplotypes that lead to diabetes susceptibility, neutral or protective are identified from this work. The significant islet autoantigens such as IA-2, GAD-67 autoantibodies and their role in causing the disorder is studied from [20]. The paper on “Genetics of type 2 diabetes mellitus” by Wyso, MCV Ng, SC Lee, T Sanke, HK Lee, JCN Chan in HKMI Vol. 6 No. 1, Mar 2000 paper [21] gave the details about the roles of insulin and counter-regulator hormones in the regulation of glucose homeostasis. The specific types of diabetes mellitus, DNA binding factors involved in the transcription in the pancreas was identified for the same paper. The paper gave an important view about the pathogenesis of type 2 diabetes mellitus. The reference titled “No diabetes association mutations in the coding region of the hepatocyte nuclear factor – 4 γ gene (HNF4G) in Japanese patients with MODY [22]” gave a very significant conclusion that mutations in transcription factors expressed in the pancreatic beta cell can cause diabetes mellitus, including the Maturity On Set Diabetes of the young and early onset autosomal–dominant type II (non insulin dependant) diabetes mellitus. The Epidemiology by the article of James H. Warram Stephen
S. Rich and Andrezaj S. Krolewske [23] gave a description of epidemiology of insulin dependent diabetes mellitus (Type I) and non-insulin dependant diabetes mellitus (Type II) including a discussion on the etiology of the disease. The reference on “Neural Networks for protein structure prediction by Burkhard Rost CUBIC Columbia University” [24] explained how neural networks have been applied to many pattern classification problems. The fuzzy logic application was studied from the Mizymoto reference from “Improvement methods of Fuzzy Controls” in the proceedings of the 3rd IFSA Congress page 60-62, 1989. The importance of fuzzy expert system, the set of rules to be used in the fuzzy expert system forming the rule base or knowledgebase was learnt from the same paper. The paper on “Making Computer think like people” IEEE spectrum, 8/1984, pp 26-32. S. Hacli, “DO WE NEED FUZZY LOGIC?” Int Journa of Man-Mach Stud, Vol. II, 1979, pp 437-445 gave a complete view on the fuzzy logic, fuzzy sets and application of the same for real time problems. The research paper on “MODEL SELECTION FOR NEURAL NETWORKS CLASSIFICATION” by Herbert K.H. Lee, Pake University [25] gave the details of developing a neural network model for the Pima Indian Diabetes Data Base and analysis of the same. The Bayesian approach to mode selection was also derived from the study and implementation is carried out with the data sample taken for the study. With all the above literature study the problem was well formulated in the diagnostic model, applied with the fuzzy logic to show risk levels of diabetic gene patterns.
and finally a mathematical and Neural Network approach is carried out to draw
different conclusions about the disease causing phenotypic parameters in
medical diagnostics. Data mining techniques are applied to highlight the
significant role played by age, obesity and hereditary factors in the cause of
diabetes mellitus.

2.1.1. LIMITATIONS IN THE LITERATURE SURVEY

- Pima Indian diabetes database was not tested with correlation coefficients to
  find the interdependence between each of the parameters.

- Pima Indian diabetes database was not grouped and analyzed by data
  mining method using naïve Bayesian classifiers, and logistic regression.

- Pima Indian diabetes database was not grouped and analyzed by data
  mining method using ID3 algorithms.

- Yule’s coefficient was not applied with reference to Pima Indian diabetes
  database.

- The importance of diabetes pedigree function to indicate the hereditary
  nature of diabetes using neural network approach was not dealt with, in the
  above surveyed papers and websites.
2.2. VARIOUS MODULES AND BLOCK DIAGRAMS

Nucleotides (A, G, C, T) -> m-RNA (TasU) -> Aminoacids -> Chain of aminoacids ->
Proteins (e.g. INSULIN, HEAMOGLOBIN) - Responsible for all Enzymes and
Hormones of a body -> Expressive part of Protein codes for a gene (i.e.
Particular expressive part of A GC T forms the gene).

2.2.1. DIAGNOSTIC MODEL OF THE PROBLEM
Module: 1

Diagnostic model of the problem defines diabetes mellitus as a pancreatic disorder and the endocrine process behind the insulin secretion and glucose metabolism are explained. The model also distinguishes the different forms of diabetes mellitus (type 1, type 2 and other forms). The genetics of type 1 and type 2 diabetes are widely described in this module. The various mutations causing sudden onset of diabetes in children and adults are also discussed in this model. Mutations may occur due to exposure to radiations and chemical agents and temperature changes in the working environment of an individual. Pictorial demonstrations are shown in the different forms of mutations.fig[2.1]

2.2.2. FUZZY-GENETIC MODEL OF THE PROBLEM
Module: 2

Applying Fuzzy-logic based model to the problem to analyze the cluster of selective genes, it is to identify the mutated and normal regions of DR, DQ
alleles and determine their susceptibility probabilities to Diabetes Mellitus. Tables and Graphs are developed to analyze and discuss the results obtained from various computer based Genome Analysis over a restricted set of available alleles of the HLA region. The module can indicate the probability of an individual inheriting a risky gene leading to susceptibility of the disease. The genes can also be protective in suppressing the disease. The fuzziness in the risk levels fig[2.2] of different diabetic genes are modelled and indicated by graphical demonstration. The importance of genetic markers in susceptibility to type 1 diabetes is outlined in this model. Finally the uncertainty in fuzzy logic model is discussed with an example. Fuzzy cluster analysis to predict the type of diabetes acquired is discussed with a sample data.

2.2.3. MATHEMATICAL MODEL OF THE PROBLEM
Module : 3

To give a mathematical study of the relationship between various factors and the total effect in the cause of the disease correlation and regression coefficients can be used. Regression is generally used to predict future values based on past values by fitting a set of points on the curve. Correlation however, is used to examine the degree to which values two or more variables behave in a similar manner. Here, they are used to find the relationship between various factors causing diabetes mellitus Yule’s coefficient is
developed to find the associations between the diabetic and non-diabetic attributes.

The following statistical coefficients have been calculated for the study from the 8 input parameters.

(1) Spearman's correlation coefficient & Karl Pearson's correlation coefficient
(2) Yule's coefficient for association of attributes
(3) Posterior probabilities of the attributes

2.2.4. NEURAL NETWORK MODEL FOR THE PROBLEM
Module: 4

The posterior probability factors of statistical model are used to build Neural network model using BPN network theory of Artificial Neural Networks. The weights of neural network are determined by training the network. In neural network procedure, training is based on learning of diabetes causing factors, which follows the necessary algorithms to highlight the significant factors causing diabetes. Here, training to detect proneness to diabetes is a mathematical minimization problem. Since the outcome is known (whether diabetes acquired or not) the supervised learning method is followed. The estimation of the posterior probability is developed to reach the activation function and to the necessary errors are back propagated to minimize the same. Let each object be associated with a d-vector x of attributes. Assume that the
sample space $X$ consists of $m$ groups of the objects. Following Duda and Hart (1973), let $w_j$ denote the fact that an object is a matter of group $j$.

Let us define

$P(w_j) = \text{prior probability group of } j; \text{ the probability that a randomly selected object belongs to group } j$;

$P(x | w_j) = \text{conditional probability distribution function for } x \text{ being a member of group } j$.

The posterior probability $P(w_j | x)$, which is the probability that object $x$ belongs to group $j$, is obtained using the Bayes rule:

$$P(w_i | x) = \frac{P(x | w_i)P(w_i)}{\sum_{j=1}^{m} P(x | w_j)}$$

For medical diagnosis, one group may be positive (e.g., diabetic) whereas the other group may be negative (e.g., not diabetic) [26]. Then the posterior probability of the first group will be the probability that the case is positive.

For two-group classification problems, one output node is sufficient. So the objective function is simplified to

$$\text{Minimize } \sum_{i=1}^{X} (y^i - a^i)^2$$

where $y^i = 1$, if pattern $i$ belongs to group 1 and $y^i = 0$ otherwise, and $a^i$ denotes the neural output for pattern $i$. 
2.2.5. DATAMINING MODEL OF THE PROBLEM

Module: 5

In module 5 Bayesian classifiers are used in the analysis of data using Bayes theorem, naive Bayesian classifier, logistic regression coefficient. ID3 algorithm is used to draw the decision trees. In the Bayesian classification from the pima Indian diabetes database sample records are fuzzified and probability of acquiring the disorder is predicted for any individual when the parameters causing the disorder are properly estimated and given. The plot of every significant factor in the cause of disease are shown by graphical demonstrations. Logistic regression is applied to bring out the role of age and hereditary factors behind the onset of diabetes. ID3 algorithm is applied to draw decision trees for the diabetes causing parameters.
Figure [2.1] Pathogenesis of Diabetes Mellitus - Type - 2

- Physical inactivity
  - High-fat diet
  - Stress
  - Ageing
- Insulin resistance
- Obesity (central or general)
- Genetic factors
- Poor pancreatic β-cell development
- Poor in utero nutrition
- Blood glucose (glucose toxicity)
- Insulin secretion
- β-cell failure
- Blood glucose (glucose toxicity)
NOTE:--
DR2, DR3, DR4, DR5, DQGENES - EXAMPLES OF VARIOUS DIABETES
CAUSING GENES
RISK LEVELS R1, R2, R3, R4: EXTENT TO WHICH THESE GENES
CAUSE DIABETES OR PRONENESS TO DIABETES GENETICALLY.