CHAPTER – IV
CHAPTER – IV
FUZZY – GENETIC MODEL

4.1. GENETIC COMPONENTS IN IDDM [46]

The MHC complex haplotype particularly genes coding for the so called class II proteins play a major role in IDDM [47] susceptibility. Normally the chance of any two siblings inheriting the same haplotype is 25%. However, if both have IDDM identify between haplotypes occur between 55 – 60% of cases.

Significance of HLA (Human Leukocyte Antigen) region is a section of a chromosome that contains several genes that are involved in how the immune system works. These genes make proteins that dot the surface of some cells in the Immune system. These proteins are important for the immune system to distinguish between its own cells and an infectious agents such as a bacteria or virus. When this system fails the immune cells attack other cells of the body (such as pancreatic cells) in a process called autoimmune reaction.

There are two major genes in the HLA region that account for 40 to 50 percent of diabetes risk that people inherit from their parents. Different versions (or alleles) of these genes can put a person at risk for or prevent them from developing Type I diabetes. One of the genes in the HLA region that plays an important role in Diabetes is called DR. People can inherit one from of DR from their mother and one from their father. It is the combination
of these two forms of the gene determines a person’s overall risk. Two forms of DR, designated, as DR3 and DR4 are present in 95 percent of Type I Diabetes and 30 percent generally inherit both DR3 and DR4. Fifty percent of general population inherits DR3 or DR4 and the inheritance differs in different ethnic groups.

Table [4.1] Alleles and Risk Level of Diabetic Genes

<table>
<thead>
<tr>
<th>Allele</th>
<th>Risk Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>DR1</td>
<td>Slight risk</td>
</tr>
<tr>
<td>DR2</td>
<td>Protective</td>
</tr>
<tr>
<td>DR3</td>
<td>Significant risk</td>
</tr>
<tr>
<td>DR4</td>
<td>Significant risk</td>
</tr>
<tr>
<td>DR5</td>
<td>Slight risk</td>
</tr>
<tr>
<td>DR6</td>
<td>Neutral / Protective</td>
</tr>
<tr>
<td>DR7</td>
<td>Protective Risk in African Descent</td>
</tr>
<tr>
<td>DR8</td>
<td>Neutral / Slight risk</td>
</tr>
<tr>
<td>DR9</td>
<td>Risk in Chinese, Japanese, Korean Descent</td>
</tr>
</tbody>
</table>

Although both DR3 and DR4 alleles put a person at risk for developing Diabetes, the two alleles cause slight differences in the disease fig [3.8] to fig [3.10].

**Diabetics who have inherited DR3 (but not DR4) develop diabetes at an older age, and tend to have antibodies against Insulin. These people are also more likely to develop thyroid autoimmune disease.**
Diabetics who have inherited DR4 (but not DR3) tend to develop diabetes earlier in life and have an immune reaction against Insulin.

Diabetics who have inherited DR3 and DR4 develop diabetes at the youngest age and have the highest levels of antibodies against Insulin.

Another gene in the HLA region called DQ also contributes to the development of Type I diabetes like DR certain versions of the DQ gene but a person at higher risk for developing the disease, while the other forms seem to be protective. To make matters worse people who inherit DR3 or DR4 also tend to inherit a form of DQ that adds a very high risk of developing diabetes. Conversely, the protective forms of DQ and DR tend to be inherited together. In fact, these DQ alleles may be responsible for the risk or prevention reported for the different DR alleles.

The Human Leukocyte Antigen (HLA) complex is located on the short arm of chromosome 6 at p21.3 (1-4). It encompasses approximately 3500 kb of DNA, and contains at least 150 genes. It is the primary region of susceptibility for type 1 diabetes, as well as other autoimmune disorders. Recent genome screens have identified the class II sub-region (i.e., HLA-DR, DQ, DP loci) as IDDM1.

The DQ locus, which is the focus of this review, consists of two tightly linked genes (i.e., DQA1 and DQB1) that encode alpha and beta glycoproteins, respectively. These molecules combine non-covalently to form functional α-β heterodimers. DQA1 and DQB1 genes are highly polymorphic. Allelic
variation is observed primarily in the second exon, which corresponds to the peptide-binding cleft. HLA-DQ and other class II molecules present extracellular antigens to helper T-cells and stimulate immune response. They have a restricted tissue distribution and are located mainly on macrophages, B cells and activated T cells.

Transcription of DQA1 and DQB1 is complex and involves cis- and trans-acting factors. Critical upstream regulatory sequences have been reported for DQA1 and DQB1. Variation in promoter sequences affects gene expression, and may also be involved in the pathogenesis of autoimmune disorders. In addition, post transcription activities appear to influence disease risk. For example, functional DQβα heterodimers can be formed from the non-covalent association of products of DQA1 and DQB1 genes in cis. Alternatively, the combined α and β glycoproteins can represent molecules encoded by genes in trans. Hybrid DQ molecules with DR or DP glycoproteins have also been observed.

4.1.1. HLA-DQ AND THE INSULIN GENE

In Caucasians, it has been demonstrated that the insulin gene region (INS), located on chromosome 11p15.5, contains the second major susceptibility locus for type 1 diabetes (i.e., IDDM2) [48]. Positive associations have been observed with a non-transcribed minisatellite region (VNTR) in the 5' flanking region. There are two common alleles; the shorter
class I allele predisposes to type 1 diabetes, while the longer class III allele appears protective. The biological plausibility of these associations may relate to the expression of insulin mRNA in the thymus. Class III alleles generate higher levels of insulin mRNA than class I alleles. These differences may contribute to a better immune tolerance for class III positive individuals by increasing the likelihood of negative selection for auto-reactive T cell clones.

The Diabetes Prediction and Prevention Project (DIPP) is designed to determine whether it is possible to delay the clinical manifestations of type 1 diabetes with nasal insulin by at least three years in high risk Finnish newborns. The Diabetes Autoimmunity Study in the Young (DAISY) is a natural history study, which is also based on newborn screening for high risk HLA class II alleles. This investigation involves families of varied ethnic backgrounds (Hispanic, Non-Hispanic white, Black and Asian). The susceptibility alleles required for inclusion were defined as DRB1*03, DRB1*04 and DQB1*0302; and DRB1*15/16 (DR2) was considered as an exclusion criteria. Since the study is in its early phases, the sensitivity and specificity of the screening have yet to be determined.

In summary, HLA-DQ [49] screening for type 1 diabetes is now being conducted in high risk families and the general population for intervention trials and natural history studies. Thus, there is a critical need to reconsider the risks, benefits, ethics, legal and social issues regarding genetic and/or autoantibody testing for type 1 diabetes. In addition, population-based risk
factor specific incidence rates are urgently needed for all ethnic groups. Translating research findings for prediction and prevention outside a research environment also requires genetic counseling and genetic education programs for type 1 diabetes [50] family members, as well as health care professionals. During the next millennium, these issues should be among the top priorities for finding for type 1 diabetes.

4.1.2. The Human Leukocyte Antigen (HLA)

The Human Leukocyte Antigen (HLA) complex is located on the short arm of chromosome 6 at p21.3 (1-4). It encompasses approximately 3500 kb of DNA, and contains at least 150 genes. It is the primary region of susceptibility for type 1 diabetes, as well as other autoimmune disorders. Recent genome screens have identified the class II sub-region (i.e., HLA-DR, DQ, DP loci) as IDDM1.

The DQ locus, which is the focus of this review, consists of two tightly linked genes (i.e., DQA1 and DQB1) that encode alpha and beta glycoproteins, respectively. These molecules combine non-covalently to form functional α-β heterodimers. DQA1 and DQB1 genes are highly polymorphic. Allelic variation is observed primarily in the second exon, which corresponds to the peptide-binding cleft. HLA-DQ and other class II molecules present extra cellular antigens to helper T-cells and stimulate immune response. They have a
restricted tissue distribution and are located mainly on macrophages, B cells and activated T cells.

Transcription of DQA1 and DQB1 is complex and involves cis- and trans-acting factors. Critical upstream regulatory sequences have been reported for DQA1 and DQB1. Variation in promoter sequences affects gene expression, and may also be involved in the pathogenesis of autoimmune disorders. In addition, post transcription activities appear to influence disease risk. For example, functional DQβα heterodimers can be formed from the non-covalent association of products of DQA1 and DQB1 genes in cis. Alternatively, the combined α and β glycoproteins can represent molecules encoded by genes in trans. Hybrid DQ molecules with DR or DP glycoproteins have also been observed.

4.1.3. HLA REGION AND DIABETES RISK

HLA genes are contained in the major histocompatibility complex (MHC) on the short arm of chromosome 6 and are classified into class II, III genes, and I. The class III genes are further divided into sub regions of DR, DQ, DN/DO and DP comprising the expressed, pseudo genes and the genes without known product. The molecules of class II are expressed uniquely on the surface of certain types of cells: macrophages active B and T lymphocytes, dendritic and endothelial cells. The HLA molecules have the function to present antigenic peptide to T lymphocytes. The recognition of antigens
attached to class II molecules by cd4+ lymphocytes leads to the activation of auxiliary T lymphocytes and the immune response. The structure at the level of the sites implied in the liaison of the antigen seems crucial for the function of the molecule.

Every gene has a number of alleles coding for the diverse molecules expressed in different individuals. The distribution of these alleles are specific for every population. The combination of alleles which are always carried by a chromosome constitute a HLA-haplotype. The polymorphism of the expressed class II molecules already considerable due to complexity of genetic level is increased due to expression of hybrid molecules. These are the molecules of trans complementation resulting from the combination in trans of a and b chains coding for the genes A and B of two different chromosomes while the inter-isotopic molecules are formed by a and b chains of two different loci (E.g., DRa –DWb). The alleles found at different loci of a haplotype are not distributed just by chance but according to certain preferred association called” linkage disequilibrium (Fig.3.5 to Fig.3.7). The genes, which are extremely close like that of, class II and with strong linkage disequilibrium is associated to a disease. With the evolution of techniques allowing finer analysis of HLA system, the markers associated to IDDM were localized more precisely. Each different amino acid sequence is given a number. For the DQ molecules both its alpha and beta chain gene are polymorphic and thus to specify a DQ
molecule one must specify both chains. For DR molecules only the DRB chain is polymorphic and thus only this chain is specified. Each number after the star indicates a specific amino acid sequence of the HLA allele (Figure 6) and the letters and first number the gene (e.g. DRB1*0401, DR B chain gene number 1, allele 0401). The alleles of different HLA genes (e.g. DRB1 and DQB1) are non-randomly associated with each other, such that with DRB1*0401, one usually finds one of three DQ alleles (e.g. DQB1*0301, DQB1*0302, DQB1*0303) rather than any one of more than forty different DQB molecules. Such non-random association of alleles of different genes on the same chromosome is termed linkage disequilibrium.

The serologic typing of class II antigens showed that frequency of antigens DR3 (DW3) and DR4 (DW4) were much more increased among IDDM subjects compared to healthy control population. The DR region was more closely associated to IDDM and the association with class I was secondary to linkage disequilibrium between B15 and DR4 and between B8 or B18 and DR3. Similarly negative association was more strongly linked to DR2 in linkage disequilibrium with B7. The table below reveals that the frequency of different HLA antigens and their associations with diseases vary with the ethnic and geographic origin. In certain populations there is positive association with DR9 (Chinese, Japanese, blacks) or DR7 (blacks). In Caucasians more than ninety percent of the IDDM patients are DR3 or
The highest risk is observed with the heterozygote subjects having DR3 / DR4 which represent thirty to fifty percent of juvenile IDDM patients. Recently it has been found on the position 57 of the DQB chain correlate with the susceptibility or resistance to IDDM. Most of the haplotypes neutral or negatively associated to IDDM carry an aspartic residue (ASP57), ("protective allele") while the haplotypes positively associated to IDDM code for noncharged ("susceptibility allele") amino acid (Ala, val or ser 57). These observations were confirmed in transgenic mouse model.

Presentation of antigenic peptides of islet auto antigens is a key process in the autoimmune response that leads to type 1 diabetes mellitus. Auto antigens like GAD65, IA- 2 play a crucial role in the development of type1 diabetes known as (LATENT AUTO IMMUNE DIABETES FOR ADULTS OR LADA). Researchers have developed intramuscular immunization with plasmid DNA expressing pancreatic islet auto antigens (GAD or insulin B Chan) to protect from type 1 diabetes. The induction of auto reactive, regulatory (D4 lymphocytes that produce IL-4 rather than deletion or "anergy" of auto aggressive T-Cells was the mechanism underlying this protection. Table [3.11] illustrates the frequency of DQA1 allele varied across ethnic groups.
4.2. LOGICAL IMPLEMENTATION

4.2.1. FUZZY LOGIC BASED ANALYSIS

EXAMPLE OF FUZZINESS [51] WITH RESPECT TO OUR PROBLEM

In our problem we proceed forward to predict the risk-levels of Type I Diabetes and risky alleles inherited. Based on the type of allele inherited by a person we categorize the risk level under the following uncertain categories. They are, HIGHLY SIGNIFICANT, SIGNIFICANT, PROTECTIVE.

Fuzzy classification & Membership validation

The fuzzy categories we have are:

- Slight risk
- Significant risk
- Highly Significant risk

Based on the type of allele acquired, the risk levels are categorized as follows:

<table>
<thead>
<tr>
<th>Slight Risk</th>
<th>Significant Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Allele Acquired</strong></td>
<td><strong>Interval</strong></td>
</tr>
<tr>
<td>DR 1</td>
<td>0.1</td>
</tr>
<tr>
<td>DR 5</td>
<td>0.1</td>
</tr>
<tr>
<td>DR 8</td>
<td>0.1</td>
</tr>
<tr>
<td>DR 1 &amp; DR 5</td>
<td>0.3</td>
</tr>
<tr>
<td>DR 1 &amp; DR 8</td>
<td>0.4</td>
</tr>
<tr>
<td>DR 5 &amp; DR 8</td>
<td>0.4</td>
</tr>
<tr>
<td>DR 1, DR 5 &amp; DR 8</td>
<td>0.55</td>
</tr>
<tr>
<td>DR 4 &amp; DR 8</td>
<td>0.6</td>
</tr>
</tbody>
</table>
Highly Significant Risk

<table>
<thead>
<tr>
<th>Allele Acquired</th>
<th>Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>DR 3 &amp; DR 4</td>
<td>0.85</td>
</tr>
<tr>
<td>DQ</td>
<td>1.0</td>
</tr>
</tbody>
</table>

MEMBERSHIP FUNCTIONS

The membership functions for fuzzy categories we have [52],

\[ M() : \text{MEMBERSHIP FUNCTION} \quad R_1 : \text{RISK LEVEL} \]

\[ M(\text{Slight Risk}) = \begin{cases} 1 & \text{if } R_1 \geq 0.1 \text{ and } R_1 \leq 0.6 \\ 0 & \text{otherwise} \end{cases} \]

\[ M(\text{Significant Risk}) = \begin{cases} 1 & \text{if } R_1 \geq 0.5 \text{ and } R_1 \leq 0.8 \\ 0 & \text{otherwise} \end{cases} \]

\[ M(\text{Very Significant}) = \begin{cases} 1 & \text{if } R_1 \geq 0.7 \text{ and } R_1 \leq 1.0 \\ 0 & \text{otherwise} \end{cases} \]

Fuzzy graphs can be drawn between risk levels and membership functions.
4.3. GRAPHICAL ANALYSIS OF RESULTS

Figure [4.1] Graphic Analysis of Results

4.3.1. DETERMINATION OF MEMBERSHIP TO FALL UNDER A PARTICULAR CATEGORY

From the fuzzy graphs we can proceed forward to calculate the possibility of a risk level to fall under a particular category. The procedure is as follows: -For example here, we take the following values:-

We are to find under which category does the risk level category of 0.55 fall into. From the graph we can find that the category 0.55 falls under both the areas of slight risk and significant risk. Let us now see exactly under what category does it exactly fit into. From the FUZZY GRAPH take the height of category 0.55 in the area of Slight Risk. Taking probability for Slight risk
category we get $0.5 / 0.7 = 0.7142$. Taking probability for significant risk category we get $0.2 / 0.7 = 0.2857$. From the above two probability values we see that category of 0.55 falls under the slight risk.

Table [4.2] Graphical Analysis of Results

TABLES AND GRAPHS

<table>
<thead>
<tr>
<th>Genetic marker</th>
<th>No IDDM relative</th>
<th>First degree relative with IDDM</th>
</tr>
</thead>
<tbody>
<tr>
<td>DQM*0302/*0201</td>
<td>1 in 25</td>
<td>1 in 4</td>
</tr>
<tr>
<td>DQM*0302/*0302</td>
<td>1 in 60</td>
<td>1 in 10</td>
</tr>
<tr>
<td>DQM*0302/*602</td>
<td>1 in 1500</td>
<td>Unknown</td>
</tr>
<tr>
<td>DQM*0302/*other</td>
<td>1 in 60</td>
<td>1 in 10</td>
</tr>
<tr>
<td>DQM*0201/*0201</td>
<td>1 in 350</td>
<td>1 in 10</td>
</tr>
<tr>
<td>DQM*0201/*other</td>
<td>1 in 400</td>
<td>1 in 20</td>
</tr>
<tr>
<td>Other</td>
<td>1 in 5000</td>
<td>1 in 40</td>
</tr>
</tbody>
</table>

4.3.2. IMPORTANCE OF GENETIC MARKERS IN SUSCEPTIBILITY TO TYPE 1 DIABETICS

Risk estimates for susceptibility to type 1 diabetes in Caucasians according to presence or absence of IDDM susceptibility genes and family history. DQB*0302 and DQB*0201 are associated with high susceptibility, while DQB*0602 is a protective gene. The risk is highest in patients who are heterozygous for both high susceptibility genes – DQB*0302 and DQB*0201 (Data from Neprom, GT, Diabetes Rev 1993; 1:93) [22].
4.4. UNCERTAINTY IN FUZZY LOGIC MODEL

Imprecision and uncertainty play a large role in the field of medicine. The field has, for this reason, become one of the most fruitful and active areas of application for the theory of evidence. With the increased volume of information available to physicians from new medical technologies, the process of classifying different sets of symptoms under a single name and determining the appropriate therapeutic actions becomes increasingly difficult. A single disease may manifest itself quite differently in different patients and at different disease stages. Further, a single symptom may be indicative of several different disease, and the presence of several disease in a single patient may disrupt the expected symptom pattern of any one of them. The best and most useful descriptions of disease entities often use linguistic terms that are irreducibly vague. Although medical knowledge concerning the symptom-disease relationship constitutes one source of imprecision and uncertainty in the diagnostic process, the knowledge concerning the state of the patient constitutes another. The knowledge provided by each of these sources carries with it varying degrees of uncertainty. The past history offered by the patient may be subjective, exaggerated, underestimated, or incomplete. These models vary in the degree to which they attempt to deal with different complicating aspects of medical diagnosis such as the relative importance of symptoms, the varied symptom patterns of different disease stages, relations between diseases themselves, and the stages of hypothesis formation, preliminary diagnosis, and
The fuzzy set framework has been utilized in several different approaches to modelling the diagnostic process. In the approach formulated by Sanchez, the physician's medical knowledge is represented as a fuzzy relation between symptoms and diseases. Thus, given the fuzzy set $A$ of the symptoms observed in the patient and the fuzzy relation $R$ representing the medical knowledge that relates the symptoms (of diabetes) in set $S$ to the diseases (categories of diabetes) in set $D$, then the fuzzy set $B$ of the possible diseases of the patient can be inferred by means of the compositional rule of inference

$$B = A \circ R$$

or

$$\mu_B(d) = \max[\min(\mu_A(s), \mu_R(s,d))]$$

for each $d \in D$. This max-min composition corresponds to the fuzzy conditional statement "If $A$ then $B$ by $R$." The membership grades of observed symptoms in fuzzy set $A$ may represent the degree of certainty of the presence of the symptom or its severity. The membership grades in fuzzy set $B$ denote the degree of certainty with which we can attach each possible diagnostic label to the patient. The fuzzy relation $R$ of medical knowledge should constitute the final diagnosis within the diagnostic process itself. These models also form the basis for computerized medical expert systems, which are usually designed to aid the physician in the diagnosis of some specified category of diseases like diabetes mellitus.
greatest relation such that given the fuzzy relation \( Q \) on the set \( P \) of patients and \( S \) of symptoms and the fuzzy relation \( T \) on the set \( P \) of patients and \( D \) of diseases, then

\[
T = Q \circ R
\]

Thus, relations \( Q \) and \( T \) may represent, respectively, the symptoms that were present and diagnoses that were consequently made for a number of known cases. Figure 6.4 summarizes the meanings and uses of fuzzy-relations \( Q, T, \) and \( R \) and fuzzy sets \( A \) and \( B \). By solving the fuzzy relation equation for \( R \), the accumulated medical experience can be used to specify the relation between symptoms and diseases that was evidenced in the previous diagnoses. The maximal solution to Eq. must be chosen for \( R \) in order to avoid arriving at a relation that is more specific than our information warrants. However, this can lead to cases in which \( R \) shows more symptom-disease association than exists in reality. Therefore, it may be necessary to interpret the results of applying relation \( R \) to a specific set of symptoms as a diagnostic hypothesis rather than as a confirmed diagnosis.

The [53] model proposes two types of relations to exist between symptoms and diseases: an occurrence relation and a conformability relation. The first provides knowledge about the tendency or frequency of appearance of a symptom when the specific disease is present; it corresponds to the question, "How often does symptoms occur with diseased? The second relation
describes the discriminating power of the symptom to confirm the presence of the disease; it corresponds to the question, "How strongly does symptom s confirm disease d?" The distinction between occurrence and conformability is useful because a symptom may be quite likely to occur with a given disease but may also commonly occur with several other diseases, therefore limiting its power as a discriminating factor among them. Another symptom, on the other hand, may be relatively rare with a given disease, but its presence may nevertheless constitute almost certain confirmation of the presence of the disease.

For this example, let \( S \) denote the crisp universal set of all symptoms, \( D \) be the crisp universal set of all diseases, and \( P \) be the crisp universal set of all patients. Let us define a fuzzy relation \( R_s \) on the set \( P \times S \) in which membership grades \( \mu_{RS}(p,s) \) (where \( p \in P, s \in S \)) indicate the degree to which the symptom \( s \) is present in patient \( p \). For instance, if \( s \) represents the symptom of increased potassium level and the normal test result range is roughly 3.5 to 5.2, then a test result of 5.2 for patient \( p \) could lead to a membership grade \( \mu_{RS}(P,s) = 0.5 \). Let us further define a fuzzy relation \( R_o \) on the universal set \( S \times D \), where \( \mu_{Rs}(s,d) \) (\( s \in S, d \in D \)) indicates the frequency of occurrence of symptom \( s \) with disease \( d \). Let \( R_c \) also be a fuzzy relation on the same universal set, where \( \mu_{RC}(s,d) \) corresponds to the degree to which symptom \( s \) confirms the presence of disease \( d \)(diabetes mellitus).
Here, we will determine the fuzzy occurrence and confirmability relations from expert medical documentation. Since this documentation usually takes the form of statements such as "Symptom s seldom occurs in disease d" or "Symptom s always indicates disease d," we assign membership grades of 1, .75, .5, .25, and 0 in fuzzy sets Ro and Rc for the linguistic terms always, often, unspecific, seldom, and never, respectively. We use a concentration operation to model the linguistic modifier very such that

$$\mu_{\text{veryA}}(x) = \mu_A^3(x)$$

Assume that the following medical documentation exists concerning the relations of symptoms $s_1, s_2,$ and $s_3$ to diseases $d_1$ and $d_2$:

- Symptom $s_1$ occurs very seldom in patients with disease $d_1$ (Type-1 diabetes).
- Symptom $s_2$ often occurs in patients with disease $d_2$ but seldom confirms the presence of disease $d_2$ (Type-2 diabetes).
- Symptom $s_2$ always occurs with disease $d_1$ and always confirms the presence of disease $d_1$: $s_2$ never occurs with disease $d_2$ and (obviously) its presence never confirms disease $d_2$.
- Symptom $s_3$ very often occurs with disease $d_2$ and often confirms the presence of $d_2$ (Type-2 diabetes).
- Symptom $s_3$ seldom occurs in patients with disease $d_1$ (Type-1 diabetes).
All missing relational pairs of symptoms and diseases are assumed to be unspecified and are given a membership grade of .5. From our medical documentation we construct the following matrices of relations $R_0$, $R_c \in S \times D$:

\[
R_0 = \begin{bmatrix}
 s_1 & s_2 & s_3 \\
 0.06 & 1 & 0.25 \\
 0.75 & 0 & 0.56 \\
\end{bmatrix}
\]

\[
R_c = \begin{bmatrix}
 s_1 & s_2 & s_3 \\
 0.5 & 1 & 0.5 \\
 0.25 & 0 & 0.75 \\
\end{bmatrix}
\]

* For the sake of simplicity, we denote the matrix representation of any relation $R$ by $R$.

Now assume that we are given a fuzzy relation $R_s$ specifying the degree of presence of symptoms $s_1$, $s_2$, and $s_3$ for three patients $p_1$, $p_2$, and $p_3$ as follows:

\[
R_s = \begin{bmatrix}
 p_1 & p_2 & p_3 \\
 0.4 & 0.6 & 0.9 \\
 0.8 & 0.9 & 0 \\
 0.7 & 0 & 1 \\
\end{bmatrix}
\]

Using relations $R_s$, $R_0$, and $R_c$, we can now calculate four different indication relations defined on the set $P \times D$ of patients and diseases. The first is the occurrence indication $R_1$ defined as

\[
R_1 = R_s \circ R_0
\]
For our example, $R_1$ is given by the following matrix:

$$
R_1 = \begin{pmatrix}
\ p_1 & d_1 & .8 \\
\ p_2 & d_2 & .9 \\
\ p_3 & .25 & .6 \\
\end{pmatrix}
$$

The confonnability indication relation $R_2$ is calculated by

$$
R_2 = R_s \circ R_c,
$$

this results in

$$
R_2 = \begin{pmatrix}
\ p_1 & d_1 & .7 \\
\ p_2 & d_2 & .9 \\
\ p_3 & .5 & .25 \\
\end{pmatrix}
$$

The nonoccurrence indication $R_3$ is defined as

$$
R_3 = R_s \circ (1 \ - R_o)
$$

and specified here by

$$
R_3 = \begin{pmatrix}
\ p_1 & d_1 & .7 \\
\ p_2 & d_2 & .6 \\
\ p_3 & .9 & .44 \\
\end{pmatrix}
$$

Finally, the nonsymptom indication $R_4$ is given by

$$
R_4 = (1 \ - R_o) \circ R_o
$$

and equals

$$
R_4 = \begin{pmatrix}
\ p_1 & d_1 & .25 \\
\ p_2 & d_2 & .25 \\
\ p_3 & .1 & .56 \\
\end{pmatrix}
$$
From these four indication relations we may draw different types of diagnostic conclusions. For instance, we may make a confirmed diagnosis of a disease \( d \) for patient \( p \) if \( \mu_{R_2}(p,d) = 1 \). Although this is not the case for any of our three patients, \( R_2 \) does seem to indicate, for instance, that disease \( d_1 \) is strongly confirmed for patient \( p_2 \). We may make an excluded diagnosis for a disease \( d \) in patient \( p \) if \( \mu_{R_3}(p,d) = 1 \) or if \( \mu_{R_4}(p,d) = 1 \). In our example, we may exclude disease \( d_1 \) as a possible diagnosis for patient \( p_3 \). Finally, we may include in our set of diagnostic hypotheses for patient \( p \) any disease \( d \) such that the inequality
\[
.5 < \max[\mu_{R_1}(p,d), \mu_{R_2}(p,d)]
\]
is satisfied. In our example, both diseases \( d_1 \) and \( d_2 \) are suitable diagnostic hypotheses for patients \( p_1 \) and \( p_2 \), whereas the only acceptable diagnostic hypothesis for patient \( p_3 \) is disease \( d_2 \). Our three types of diagnostic results, therefore, seem to point to the presence of disease \( d_2 \) in patient \( p_3 \) and disease \( d_1 \) in patient \( p_2 \), whereas the symptoms of patient \( p_1 \) do not strongly resemble the symptom pattern of either disease \( d_1 \), or \( d_2 \) alone.

The actual CADIAG-2 system incorporates relations not only between symptoms and diseases but also between diseases themselves, between symptoms themselves, and between combinations of symptoms and diseases.

An alternative model of medical diagnosis proposed by [53] combines the use of fuzzy sets and belief functions to handle two different sources of
uncertainty in the diagnostic process. In this model, diagnostic entities or
disease labels are fuzzy; each diagnostic label describes a fuzzy set $A$ defined
on the universal set $A'$ of patients that represents the class of patients with the
disease. Complete evidence concerning the state of the patient in question
would allow the physician to assign a membership grade for the patient in each
of the possible fuzzy sets corresponding to the diagnostic labels. However,
because the physician may be ignorant of the existence or importance of certain
key pieces of information or because he or she may be unable to obtain the
information either at all or to a sufficient degree of precision, only incomplete
and uncertain information regarding the patient's state is available. Let $X$
denote all the available information concerning the patients, $Y$ be all the
unavailable information, and $X(p) = x$, $Y(p) = y$ be the available and
unavailable information, respectively, that specifically concerns patient $p$. Then
the membership grade of patient $p$ in any diagnostic class $A$ depends on the
values of $x$ and $v$ such that

$$
\mu_A(p) = f(x, y)
$$

where $f$ is some appropriate function from $X \times Y$ to $[0, 1]$. Although the value
of $x$ is given, the physician has only a degree of belief concerning the value of
$y$. Therefore, the physician can, at best, express only a degree of belief, $\text{Bel}(A)$,
indicating the expectation concerning the patient's grade of membership in each
fuzzy diagnostic category $A$. A generalization of the probabilistic Bayes'
The theorem of inference to the theory of evidence can then be used to calculate a degree of belief that the patient belongs to any particular diagnostic category given the degree of belief relative to the observation of symptoms indicative of that specific pathology. Through the use of both fuzzy sets and degrees of belief, this model deals with two basic sources of uncertainty in the diagnostic process: the vagueness inherent in the diagnostic labels themselves and the limited accuracy and completeness of information or evidence concerning the status of the patient.

Another alternative approach to modeling the medical diagnostic process utilizes fuzzy cluster analysis. This type of technique is used by Fordon and Tfezdek [1979], and Esogbue and Elder [53a.]. The technique [53] of clustering examines the elements of some universal set and groups them according to similarity. Thus, elements grouped in one cluster are similar to each other and dissimilar to the members of other clusters. Since the boundaries of these clusters are not precisely defined, each cluster is, in effect, a fuzzy set in which the grade of membership of any element indicates the similarity of that element to other members of the cluster. Any particular element can, of course, belong with varying degrees to several different clusters.
4.5. FUZZY CLUSTER ANALYSIS FOR TYPES OF DIABETES

Models of medical diagnosis that use cluster analysis usually perform a clustering algorithm on the set of patients by examining the similarity of the presence and severity of symptom patterns exhibited by each. (The severity of the symptoms present can be designated with degrees of membership in fuzzy sets representing each symptom category). Often the similarity measure is computed between the symptoms of the patient in question and the symptoms of a patient possessing the prototypical symptom pattern for each possible disease. The patient to be diagnosed is then clustered to varying degrees with the prototypical patients whose symptoms are most similar. The most likely diagnostic candidates are those disease clusters in which the patient's degree of membership is the greatest.

Several different methods of fuzzy clustering exist. One group of common methods uses some form of distance measure to determine the similarity between observed attributes (symptoms) and those present in the existing diagnostic clusters. We use a simplified adaptation of the method employed by Esogbue and Elder [53] to illustrate this technique.

Let us assume that we are given a patient $x$ who displays the symptoms $s_1$, $s_2$, $s_3$, and $s_4$ at the levels of severity given by the following vector:

$$X = \begin{bmatrix} s_1 & s_2 & s_3 & s_4 \\ 1 & .7 & .4 & .6 \end{bmatrix}$$
Let $u_x(s_i) \in [0, 1]$ denote the grade of membership in the fuzzy set characterizing patient $x$ and defined on the set $S = \{ s_1, s_2, s_3, s_4 \}$, which indicates the severity level of the symptom $s_i$ for the patient. We must determine a diagnosis for this patient among three possible diseases $d_1$ (type-1), $d_2$ (type-2), and $d_3$ (other-- types). Each of these diseases is described by a matrix giving the upper and lower bounds of the normal range of severity of each of the four symptoms that can be expected in a patient with the disease. The diseases $d_1$, $d_2$, and $d_3$ are described in this way by the matrices

$$

d_1 =
\begin{array}{cccc}
\text{Lower} \\
0 & .6 & .5 & 0 \\
\text{Upper} \\
.2 & 1 & .7 & 0
\end{array}
$$

$$

d_2 =
\begin{array}{cccc}
\text{Lower} \\
0 & .9 & .3 & .2 \\
\text{Upper} \\
.2 & 1 & 1 & .4
\end{array}
$$

$$

d_3 =
\begin{array}{cccc}
\text{Lower} \\
0 & 0 & .7 & 0 \\
\text{Upper} \\
.3 & 0 & .9 & 0
\end{array}
$$

Let $\mu_{ji} (s_i) \in [0,1]$ denote the lower bound of the symptom $i$ for disease $j$, and let $\mu_{ju} (s_i) \in [0,1]$ denote the upper bound of the fuzzy symptom $i$ for disease $j$. We further define a fuzzy relation $W$ on the set of symptoms and diseases that specifies the pertinence or importance of symptom $s$ in the diagnosis of disease $d$. The relation $W$ of these weights of relevance is given by
Let $\mu_w(s_i,d_j)$ denote the weight of symptom $s_i$ for disease $d_j$. In order to diagnose the patient $x$, we use a clustering technique to determine to which diagnostic cluster (as specified by matrices $d_1$, $d_2$, and $d_3$) the patient is most similar. This clustering is performed by computing a similarity measure between the patient's symptoms and those typical of each disease $d_j$. To compute this similarity, we use a distance measure based on the Minkowski distance that is appropriately modified; it is given by the formula

$$d_p(D_j, x) = \left[ \sum_{i=1}^{m} \| \mu_w(s_i, d_j) \| (\| \mu_{d_j}(s_i) - \mu_x(s_i) \|)^p + \sum_{i=1}^{m} \| \mu_w(s_i, d_j) \| (\| \mu_{d_j}(s_i) - \mu_x(s_i) \|)^p \right]^{\frac{1}{p}}$$

where,

$$A_j = \{i/|\mu_x(s_i) < \mu_{d_j}(s_i), 1 \leq i \leq m\}$$

$$B_j = \{i/|\mu_x(s_i) < \mu_{d_j}(s_i), 1 \leq i \leq m\}$$

where $m$ equals the total number of symptoms. For this example, we give a value of $2$ to the parameter $p$ of the distance measure, thus creating a modified Euclidean metric. We use Eq. (6.7) with $p = 2$ to calculate the similarity between patient $x$ and diseases $d_1$(type-1 diabetes), $d_2$(type-2 diabetes), and $d_3$(other-type diabetes) as follows:
\[ D_2 = (d_1, x) = \left[ (0.7) (0.5 - 0.4) \right]^2 + \left[ (0.9) (0.6 - 0.6) \right]^2 \right]^{1/2} = 0.54 \]
\[ D_2 = (d_2, x) = \left[ (0.6) (0.9 - 0.7) \right]^2 + \left[ (0.8) (0.6 - 0.6) \right]^2 \right]^{1/2} = 0.19 \]
\[ D_3 = (d_2, x) = \left[ (0.9) (0.7 - 0.4) \right]^2 + \left[ (0.3) (0.7 - 0.6) \right]^2 \right]^{1/2} = 0.39 \]

The most likely disease candidate is the one for which the similarity measure attains the minimum value; in this case, the patient's symptoms are most similar to those typical of disease \( d_2 \) (type-2 diabetes).

4.6. RESULTS AND DISCUSSIONS

The application of computer aided logistics to Human Genome and prediction of diseases i.e. Diabetes Mellitus is tried on a small sphere using study of Human Leukocyte Antigen region. The significance of HLA-DR, HLA-DQ, HNF-4\( \gamma \) alleles are carried out. The nature of Diabetes is complicated, with different genetic Table[4.4] to Table[4.5] and environmental factors. But such a prediction on the genetic patterns would definitely give a probability for an individual to genetically anticipate the disease and go in for future technologies like stem cell transplanting etc to predispose the probability. Diabetics who have inherited DR3 (but not DR4) develop diabetes at an older age, and tend to have antibodies against Insulin. These people are also more likely to develop thyroid autoimmune disease. Position no 57 of the HLA-DQB CHAIN plays an integral role in the susceptibility to IDDM. An MHC CLASS-2 molecule with an aspartic acid residue at position no 57 is able to form a salt bridge between the aspartic acid residue and an arginine residue.
An MHC class II molecule from a type 1 diabetic has a substituted amino acid at position 57. The photo above shows alanine (yellow) at position 57. The presence of alanine at this position disrupts the formation of the salt bridge and is associated with increased susceptibility to IDDM.

Diabetics who have inherited DR4 (but not DR3) tend to develop diabetes earlier in life and have an immune reaction against Insulin. Diabetics who have inherited DR3 and DR4 develops diabetes at the youngest age and has the highest levels of antibodies against Insulin.

DIABETICS WHO HAVE INHERITED DRB1*0401,0405,0402,DRB1*0301, AND DQB1*0302 INDICATES HIGHEST GENETIC RISK PROBABILITIES. RESEARCH COULDN'T IDENTIFY ALL THE GENE SEQUENCES THAT PUT A PERSON AT DEFINITE RISK FOR TYPE I OR TYPE 2 DIABETES. Even if they did know all THE RISKY GENES, it is found that people with low risk genes (DR2, DR5 or long VNTR regions in chromosome 6) can still develop Diabetes and so a careful analysis of environmental factors like stress, family habits etc are to be considered along with genetic testing.
Table [4.3] Importance of Genetic Markers in Susceptibility to Type 1 Diabetics

<table>
<thead>
<tr>
<th>Genetic marker</th>
<th>No IDDM relative</th>
<th>First degree relative with IDDM</th>
</tr>
</thead>
<tbody>
<tr>
<td>DQM*0302/*0201</td>
<td>1 in 25</td>
<td>1 in 4</td>
</tr>
<tr>
<td>DQM*0302/*0302</td>
<td>1 in 60</td>
<td>1 in 10</td>
</tr>
<tr>
<td>DQM*0302/*0302</td>
<td>1 in 1500</td>
<td>Unknown</td>
</tr>
<tr>
<td>DQM*0201/*0201</td>
<td>1 in 350</td>
<td>1 in 10</td>
</tr>
<tr>
<td>DQM*0201/*other</td>
<td>1 in 400</td>
<td>1 in 20</td>
</tr>
<tr>
<td>Other</td>
<td>1 in 5000</td>
<td>1 in 40</td>
</tr>
</tbody>
</table>

Risk estimates for susceptibility to type 1 diabetes in Caucasians according fig[4.2] to fig[4.5] to presence or absence of IDDM susceptibility genes and family history. DQB*0302 and DQB*0201 are associated with high susceptibility, while DQB*0602 is a protective gene. The risk is highest in patients who are heterozygous for both high susceptibility genes – DQB*0302 and DQB*0201. (Data from Neprom, GT, Diabetes Rev 1993; 1:93) [22].
SIGNIFICANCE OF GRAPHICAL REPRESENTATIONS

**X AXIS**: TYPE OF INHERITED DIABETES ALLELE (WITHOUT IDDM ALLELE FROM ANCESTORS)

**Y AXIS**: PERCENTAGE OF RISK UPON INHERITING THE ALLELE FOLLOWING THE DISCUSSION OF DIAGNOSTIC AND FUZZY-GENETIC MODEL ARE SHOWN AS A CHARTS.

**SIGNIFICANCE OF THE GRAPHICAL PLOT**:

- PEOPLE OF DIFFERENT LEVELS OF INHERITED DIABETIC ALLELES CAN BE CAUTIONED ITS EXTENT OF PRONENESS TO THE DISEASE.

- LEVEL OF RISK WITH IDDM ANCESTORS AND WITH NO IDDM ANCESTORS ARE PLOTTED FOR COMPARISON.

- IF ANCESTORS ARE WITH IDDM ALLELES ANTICIPATION OF DIABETIC RISK ARE SHOWN.

- UPON INHERITANCE LEVEL OF RISK CAN BE ANTICIPATED.
Figure [4.3]

GENETIC MARKERS AND RISK LEVELS OF DIABETES (WITH IDDM RELATIVES)

SIGNIFICANCE OF GRAPHICAL REPRESENTATIONS

X AXIS: TYPE OF INHERITED DIABETES ALLELE (WITH IDDM ALLELE FROM ANCESTORS)

Y AXIS: PERCENTAGE OF RISK UPON INHERITING THE ALLELE FOLLOWING THE DISCUSSIONS OF DIAGNOSTIC AND FUZZY-GENETIC MODEL ARE SHOWN AS A CHARTS.

SIGNIFICANCE OF THE GRAPHICAL PLOT:

- PEOPLE OF DIFFERENT LEVELS OF INHERITED DIABETIC ALLELES CAN BE CAUTIONED OF ITS EXTENT OF PRONENESS TO THE DISEASE

- LEVEL OF RISK WITH IDDM ANCESTORS AND WITH NO IDDM ANCESTORS ARE PLOTTED FOR COMPARISON.

- IF ANCESTORS ARE WITH IDDM ALLELES ANTICIPATION OF DIABETIC RISK ARE SHOWN.

- UPON INHERITANCE LEVEL OF RISK CAN BE ANTICIPATED.
Figure [4.4]

GENETIC MARKERS AND RISK LEVELS OF DIABETES (WITH IDDM RELATIVES)

SIGNIFICANCE OF GRAPHICAL REPRESENTATIONS

**X AXIS:** TYPE OF INHERITED DIABETES ALLELE (WITH IDDM ALLELE FROM ANCESTORS)

**Y AXIS:** PERCENTAGE OF RISK UPON INHERITING THE ALLELE FOLLOWING THE DISCUSSIONS OF DIAGNOSTIC AND FUZZY-GENETIC MODEL ARE SHOWN AS A CHARTS.

SIGNIFICANCE OF THE GRAPHICAL PLOT:

- PEOPLE OF DIFFERENT LEVELS OF INHERITED DIABETIC ALLELES CAN BE CAUTIONED OF ITS EXTENT OF PRONENESS TO THE DISEASE

- LEVEL OF RISK WITH IDDM ANCESTORS AND WITH NO IDDM ANCESTORS ARE PLOTTED FOR COMPARISON.

- IF ANCESTORS ARE WITH IDDM ALLELES ANTICIPATION OF DIABETIC RISK ARE SHOWN.

- UPON INHERITANCE LEVEL OF RISK CAN BE ANTICIPATED.
Figure [4.5]

SIGNIFICANCE OF GRAPHICAL REPRESENTATIONS

**X AXIS** : TYPE OF INHERITED ALLELE (WITHOUT IDDM ALLELE FROM ANCESTORS)

**Y AXIS** : PERCENTAGE OF RISK UPON INHERITING THE ALLELE FOLLOWING THE DISCUSSIONS OF DIAGNOSTIC AND FUZZY - GENETIC MODEL ARE SHOWN AS A CHARTS.

SIGNIFICANCE OF THE GRAPHICAL PLOT

- PEOPLE OF DIFFERENT LEVELS OF INHERITED DIABETIC ALLELES CAN BE CAUTIONED OF ITS EXTENT OF PRONENESS TO THE DISEASE.

- LEVEL OF RISK WITH IDDM ANCESTORS AND WITH NO IDDM ANCESTORS ARE PLOTTED FOR COMPARISON.

- IF ANCESTORS ARE WITH IDDM ALLELES ANTICIPATION OF DIABETIC RISK ARE SHOWN.

- UPON INHERITANCE LEVEL OF RISK CAN BE ANTICIPATED.
Table [4.4] Diabetes Risk by HLA DRB, DQA and DQB Haplotypes

<table>
<thead>
<tr>
<th>RISK</th>
<th>DRB1</th>
<th>DQA1</th>
<th>DQB1</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIGH</td>
<td>0401,0405,0402 (DR4)</td>
<td>0301</td>
<td>0302</td>
</tr>
<tr>
<td></td>
<td>0301</td>
<td>0501</td>
<td></td>
</tr>
<tr>
<td>MODERATE</td>
<td>0401</td>
<td>0301</td>
<td>0301</td>
</tr>
<tr>
<td></td>
<td>0401</td>
<td>0301</td>
<td>0303</td>
</tr>
<tr>
<td></td>
<td>0403</td>
<td>0301</td>
<td>0302</td>
</tr>
<tr>
<td></td>
<td>0101</td>
<td>0101</td>
<td>0501</td>
</tr>
<tr>
<td></td>
<td>1601</td>
<td>0102</td>
<td>0502</td>
</tr>
<tr>
<td>LOW</td>
<td>1101</td>
<td>0501</td>
<td>0301</td>
</tr>
<tr>
<td>PROTECTIVE</td>
<td>1501</td>
<td>0102</td>
<td>0602</td>
</tr>
<tr>
<td></td>
<td>0701</td>
<td>0201</td>
<td>0303</td>
</tr>
<tr>
<td></td>
<td>1401</td>
<td>0101</td>
<td>0503</td>
</tr>
</tbody>
</table>
Table [4.5] HLA Terminology

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Allele</strong></td>
<td></td>
<td>$\text{DRB1}^*!0401$</td>
</tr>
<tr>
<td><strong>Haplotype</strong></td>
<td></td>
<td>$\text{DRB1}^<em>!0401 - \text{DQBI}^</em>!0302$</td>
</tr>
<tr>
<td><strong>One Chromosome</strong></td>
<td></td>
<td>$\text{DRB1}^<em>!0401 - \text{DQBI}^</em>!0302$</td>
</tr>
<tr>
<td><strong>Genotype</strong></td>
<td></td>
<td>$\text{DRB1}^<em>!0301 - \text{DQBI}^</em>!0201$</td>
</tr>
<tr>
<td><strong>Two Chromosomes</strong></td>
<td></td>
<td>$\text{DRB1}^<em>!0301 - \text{DQBI}^</em>!0201$</td>
</tr>
</tbody>
</table>