Application of Discrimination and Classification on Diabetes Mellitus Data

Michael Asamoah-Boaheng
School of Graduate Studies, Research and Innovation
Box KS 854, Kumasi Polytechnic
Kumasi, Ghana
Email: asboaheng@yahoo.com

Abstract

The assignment/allocation of individuals/observations to the various known groups with known mean vectors and distinguishing characteristics has been a major concern for years and several attempts have been made at deriving parsimonious rules that address this hurdle. In this study, Fisher’s Linear Discriminant Function (FLDF) was derived to provide maximum separation between Type 2 and Type 1 diabetes patients based on identified risk factors. The assumptions of FLDF were achieved by BoxMtest of equality of covariance matrices. A seven variate data on 620 diabetes patients obtained from Komfo Anokye Teaching Hospital (KATH) diabetes ward was obtained and used for data analyses. The derived FLDF was used to reclassify the original observation to obtain the discriminant scores from the functions and 85.3 percent correct classification was achieved. Also 84.8 percent of the cross validated grouped cases were correctly classified into either being a Type 2 or Type 1 diabetes patient group. Patients age as well as their BMI were identified to be the two major contributing variables in classifying a patient as a type1 or type 2 diabetes.

Keywords: Fisher Linear Discriminant Function, Diabetes Patients, Covariance Matrices, Cross Validation.

1.0 Introduction

Discriminant analysis is a multivariate approach for identifying the features that separate known groups or populations. In other words, discrimination is a multivariate technique concerned with separating distinct sets of observations and it is exploratory in nature. [4]. Discriminant analysis as a topic in Multivariate Statistical Analysis has attracted much research interest over the years, with the evaluation of Discriminant Functions when the covariances matrices are equal and unequal. This study is therefore aimed at using the classical method of discrimination in classifying diabetic patients as either type 1 or type 2 based on some identified anthropometric features and other factors. The problem of discrimination was first initiated by [2] in which equal covariance matrices were assumed with or without normality assumption. Fisher's approach to classification with two populations was based on arriving at a linear classification function that gave maximum separation between groups without assuming normality. Several investigations mainly with respect to multidimensional normal populations with common and unequal covariance matrices have been carried out by other authors. [6] considered the robustness of LDF under three specific distributions and the case of independent variables. These distributions were considered to be non-normal and were generated from the normal distributions using the Johnson system of transformations (i.e. log normal, inverse hyperbolic sine normal and logit normal distribution). They observed considerable decline in performance of the LDF (the log normal distribution used had extremely large skewness and kurtosis). Based on their results, Fisher's LDF was greatly affected by non-normality in the population. They concluded that, the use of Fisher's LDF under non-normality contamination situations could be badly misleading and recommended that the data be transformed to approximate normality prior to the use of the LDF.

Departures from the assumptions of linear discriminant function analysis were explored by [5], where the effects of unequal covariances on the linear discriminant method were studied. In spite of theoretical evidence supporting the use of the QDF when covariances are heterogeneous, its actual employment has been sporadic because there are unanswered questions regarding its performance in the practical situation where the discriminant function must be constructed using training samples that do not satisfy the classical assumption of the model. [8] investigated into the application of discrimination and classification on poultry feeds data.
They employed Fisher’s linear discriminant function for providing maximum separation between the two groups of eggs of which the chicken were fed with different combinations of feeds. However they proposed a linear discriminant function for the classification of eggs based on the size and cholesterol level. The function gave a good prediction based on the estimated values obtained from the Apparent Error Rates (APER) and Absolute Error Rate (AER).

[7] applied discriminant analysis in differentiating between the signal patterns of healthy subjects and those of individuals with specific heart conditions based on diagnosis of ECG signals. An approach for classifying multivariate ECG signals based on discriminant and wavelet analyses was proposed. [3] studied the variable selection criterion for linear discriminant rule and its optimality in high dimensional and large sample data. They suggested that, a new variable selection procedure called Misclassification Error Criterion (MEC) for linear discriminant rule for high dimensional data set be set up. Their study found that the MEC not only asymptotically decomposes into 'fitting' and 'penalty' terms but also possesses an asymptotic optimality in the sense that MEC achieves the smallest possible conditional probability of misclassification in candidate variable sets. After the simulation studies, the study discovered that MEC has good performances in the sense of selecting the true variable sets.

Predicting hospitalisation of patients with diabetes Mellitus; an application of the Bayesian discriminant analysis was studied by [1]. The main objective of his study was to develop and test a Bayesian discrimination model for the purpose of identifying both the personal and the healthcare system characteristics predictive of hospitalisation for the treatment of patients with diabetes Mellitus or commonly observed comorbidities associated with the disease. The model was then tested by using a logit regression technique in order to estimate the probability of one or more hospitalisation events among patients with diabetes. Claims data extracted from the Hawaii Medical Service Association (HMSA) Private Business Claims (PBS) files for the 1995 calendar year was used. The model was able to correctly classify 90 percent of the observations. The study also found that multivariate discriminant analysis using a logit regression model successfully identifies important explanatory variables predictive of hospitalisation and as well as assigns patients into 1 of 2 mutually exclusive classes.

2.0 Materials and Methodology

2.1 Data Used

A seven variate data set consisting of 620 diabetes patients either type 2 or type 1 diabetes were obtained from KATH, in Ghana and was used for data analysis. The seven measured variables included their Age, Weight (Wt), Height (Ht), Systolic Blood Pressure (BPS), Diastolic Blood Pressure (DPS), Fasten Blood Sugar (FBS) and Body Mass Index (BMI).

2.2 Discrimination and Classification of Two Populations

Let \( f_1(x) \) and \( f_2(x) \) denote the probability density function associated with a single vector random variable \( X \) for the populations \( \pi_1 \) and \( \pi_2 \) respectively. Considering an observed value \( X = (x_1, \ldots, x_p)^T \), we assign a vector \( X \) to either population \( \pi_1 \) or \( \pi_2 \). Let \( \Omega \) be the set of collection of all possible outcomes of \( X \), hence, the partition of the sample space is given as \( \Omega = R_1UR_2 \) where \( R_1 \) is the subspace of outcomes which we classify as belonging to population \( \pi_1 \) and \( R_2 = \Omega - R_1 \) the subspace of outcomes classified as belonging to \( \pi_2 \). Therefore the conditional probability of classifying an object as belonging to \( \pi_j \) when it really comes from \( \pi_i \) becomes:

\[
P(i \mid j) = P(X \in R_j \mid X \in \pi_i) = \int_{R_j} f_i(x)dx, \quad \forall \ i, j \quad i \neq j
\]  

(1)

The conditional probabilities can also be obtained for \( i = j \) when \( i, j = 1, 2 \).

Let \( P_i = P(X \in \pi_i), \ i = 1, 2 \) be the prior probability of \( \pi_i \) where \( P_1 + P_2 = 1 \). The overall probabilities of correctly and incorrectly classifying observations are:
\[ P(\text{object is correctly classified as } \pi_i) = P(X \in R_i \mid X \in \pi_i) = P(X \in \pi_i) = P(i \mid i) p_i \text{ where } i = 1, 2. \]

\[ P(\text{object is misclassified as } \pi_j) = P(X \in R_j \mid X \in \pi_j) = P(X \in \pi_j) = P(i \mid j) p_j \text{ where } i \neq j. \]

### 2.3 Cost of Misclassification

Let \( c(i \mid j) \) denote the cost of classifying an object/observation into \( \pi_i \) when actually belongs to \( \pi_j \). Where the Expected Cost of Misclassification (ECM) is derived as:

\[ ECM = c(2 \mid 1) P(2 \mid 1) p_1 + c(1 \mid 2) P(1 \mid 2) p_2 \]  
(2)

With \( p_1 \) and \( p_2 \) being the prior probabilities for the two populations. The two regions \( R_1 \) and \( R_2 \) below are used to minimized the expected cost of misclassification.

\[ R_1 = \left\{ x \in \Omega : \frac{f_1(x)}{f_2(x)} \geq \left( \frac{c(1 \mid 2)}{c(2 \mid 1)} \right) \left( \frac{p_2}{p_1} \right) \right\} \]  
(3)

\[ R_2 = \left\{ x \in \Omega : \frac{f_1(x)}{f_2(x)} < \left( \frac{c(1 \mid 2)}{c(2 \mid 1)} \right) \left( \frac{p_2}{p_1} \right) \right\} \]  
(4)

[4].

### 2.4 Classification with Two Multivariate Normal Populations when \( \Sigma_1 = \Sigma_2 \)

The density function of \( X' = (x_1, x_2, \ldots, x_p) \) for the two populations \( \pi_1 \) and \( \pi_2 \) is given by

\[ f_i(x) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp \left( -\frac{1}{2} (x - \mu_i)' \Sigma^{-1} (x - \mu_i) \right) \]

If the population parameters \( \mu_1, \mu_2 \) and \( \Sigma \) are known, then after cancellation the allocation rule after minimising the Expected Cost of Misclassification (ECM) becomes

Allocate \( x \) to \( \pi_1 \) if

\[ (\mu_1 - \mu_2)' \Sigma^{-1} x - \frac{1}{2} (\mu_1 - \mu_2)' \Sigma^{-1} (\mu_1 + \mu_2) \geq \ln \left( \frac{c(1 \mid 2)}{c(2 \mid 1)} \right) \left( \frac{p_2}{p_1} \right) \]  
(5)

[4].

### 2.5 Inferential Procedures in Discriminant Analysis

Several inferential procedures exists in discriminant function analysis. The basic ones are discussed here.

#### 2.5.1 Test for \( H_0: \mu_1 = \mu_2 \) when \( \Sigma_1 = \Sigma_2 \) using Hotelling’s \( T^2 \)-test

We assume that two independent random samples \( y_{11}, y_{12}, \ldots, y_{1n_1} \) and \( y_{21}, y_{22}, \ldots, y_{2n_2} \) are drawn from \( N_p(\mu_1, \Sigma_1) \) and \( N_p(\mu_2, \Sigma_2) \) where \( \Sigma_1 \) and \( \Sigma_2 \) are known. In order to obtain a \( T^2 \) test we assume that \( \Sigma_1 = \Sigma_2 = \Sigma \). From the samples, we calculate \( \bar{y}_1, \bar{y}_2, W_1 = (n_1 - 1)S_1 \) and \( W_2 = (n_2 - 1)S_2 \).

A pooled estimator of the covariance matrix is calculated as \( S_p = \frac{W_1 + W_2}{n_1 + n_2 - 2} \) for which \( E(S_p) = \Sigma \) hence in testing the equality of the mean vectors we use the test statistics

\[ T^2 = \frac{n_1 n_2}{n_1 + n_2} (\bar{y}_1 - \bar{y}_2)' S_p^{-1} (\bar{y}_1 - \bar{y}_2) \]  
(6)

Which is distributed as \( T^2 \sim F_{p, n_1 + n_2 - 2} \) when \( H_0 \) is true. We reject \( H_0 \) if \( T^2 \geq T^2_{p, n_1 + n_2 - 2} \).
2.5.2 Wilks Likelihood Ratio Test
If \( y_{ij}, i=1,2,\ldots, g, \quad j=1,2,\ldots,n \) are independently observed from \( N_p(\mu_i, \Sigma_i) \), then the likelihood ratio test statistics for \( H_0 : \mu_1 = \mu_2 = \ldots = \mu_g \) can be expressed as

\[
\Lambda = \frac{|E|}{|E + H|} \tag{7}
\]

Where \( H = n \sum_{i=1}^{g} (\overline{y}_i - \overline{y})(\overline{y}_i - \overline{y})' \), \( E = \sum_{i=1}^{g} \sum_{j=1}^{n} (y_{ij} - \overline{y}_i)(y_{ij} - \overline{y}_i)' \)

The test statistics is distributed as the Wilk’s \( \Lambda \)-distribution. We reject \( H_0 \) if \( \Lambda \leq \Lambda_{\alpha,p,\nu_H,\nu_E} \). \( p, \nu_H \) and \( \nu_E \) are the dimensions and degrees of freedom for hypothesis and error respectively.

2.5.3 Box’s M-Test
For a one way MANOVA with \( g \) groups \( (g \geq 2) \) the assumption of equality of covariance matrices can be stated as a hypothesis to be tested: \( H_0 : \Sigma_1 = \Sigma_2 = \ldots = \Sigma_g \) Versus \( H_1 : \) at least two \( \Sigma_i \)'s are unequal. Define

\[
W_i = \sum_{j=1}^{n} (y_{ij} - \overline{y}_i)(y_{ij} - \overline{y}_i)' \quad \text{and} \quad M = \left| \begin{array}{c|c|c} S_1 & & \\ \hline & S_2 & \\ \hline & & S_g \\ \end{array} \right|^{\nu_i/2} \left( \sum_{i=1}^{g} \nu_i / 2 \right) \tag{8}
\]

where \( \nu_i = n_i - 1 \), \( S_i = W_i / \nu_i \) is the unbiased sample covariance matrix and \( S_{pl} = \sum_{i=1}^{g} \nu_i S_i / \sum_{i=1}^{g} \nu_i = E / \nu_E \).

The statistics

\[
u = -2(1 - c_1) \ln M \tag{9}
\]

Has an approximated \( \chi^2 \) distribution with \( \frac{1}{2} (k-1)p(p+1) \) degrees of freedom where

\[
c_1 = \left[ \sum_{i=1}^{g} \frac{1}{\nu_i} - \frac{1}{\sum_{i=1}^{g} \nu_i} \left( \frac{2p^2 + 3p - 1}{6(p+1)(k-1)} \right) \right] \quad \text{we reject } H_0 \text{ if } \nu > \chi^2_{\alpha}
\]

2.6 Error Rate Estimation
The performance of any classification procedure is based on the error rates or misclassification probabilities.

2.6.1 Cross Validation
Let \( n_{1M}^{CV} \) and \( n_{2M}^{CV} \) denote the number of left out observations misclassified in group 1 and 2 respectively and it’s given by

\[
CV = \frac{n_{1M}^{CV} + n_{2M}^{CV}}{n_1 + n_2} \tag{10}
\]

3.0 Results and Discussion
First and foremost, Box M test of equality of covariance matrix of the two groups (Type 1 and Type 2) diabetic patients were tested. Table 1 shows the test of homogeneity of the two covariance matrices for Type 1 and Type 2 diabetic patients groups. From the table the log determinant values for the two groups as well as the pooled within groups were obtained. The three log determinant values as observed from Table 1 are almost the same indicating that the covariance matrices of the two diabetic patient groups are equal. The \( P \)-value of 0.350 which is greater than the significant level \( (\alpha=0.05) \) indicates that the two covariance matrices are equal (i.e. \( H_0 : \Sigma_1 = \Sigma_2 \)) and hence the data do not differ significantly from multivariate normal distribution.
3.1 Test of Equality of the Two Mean Vectors

Hotelling $T^2$ was then used to test for the equality of the mean vectors for Type 2 and Type 1 diabetic patients. The hypothesis tested for equality of the two mean vectors were:

$H_0 : \mu_1 = \mu_2$ Vs. $H_0 : \mu_1 \neq \mu_2$.

And the mean vectors for Type 1 and type 2 are:

\[
\begin{bmatrix}
\text{Age} \\
\text{Wt} \\
\text{Ht} \\
\text{BPS} \\
\text{DPS} \\
\text{FBS} \\
\text{BMI}
\end{bmatrix} = \begin{bmatrix}
29.98 \\
67.22 \\
1.64 \\
130.51 \\
80.02 \\
11.11 \\
25.06
\end{bmatrix}
\]

and

\[
\begin{bmatrix}
\text{Age} \\
\text{Wt} \\
\text{Ht} \\
\text{BPS} \\
\text{DPS} \\
\text{FBS} \\
\text{BMI}
\end{bmatrix} = \begin{bmatrix}
57.22 \\
67.61 \\
1.64 \\
134.90 \\
81.47 \\
8.93 \\
25.13
\end{bmatrix}
\]

respectively.

From Table 2 of test for equality of the two mean vectors, the $p$-value = 0.000 is less than significant level ($\alpha=0.05$), hence we reject the null hypothesis ($H_0$) and conclude that the two mean vectors of the diabetes patient groups are not equal.

Hence the pooled within group covariance matrix as well as the bivariate correlation coefficients are computed and shown in Table 3. As already indicated, the pooled within group covariance matrix satisfies one of the assumptions of Linear discriminant function and the bivariate correlation coefficients detects potential problems with multicollinearity. From the correlation table, it is clear that, none of the bivariate correlations between two of the measured variables were even closer to 0.80. This means that, multicollinearity was not observed among any two of the seven independent variables.

Next was to derive the canonical discriminant function for providing maximum separation between types 1 and 2 diabetic patients based on the identified seven independent variables. The eigenvalues table (i.e. Table 4) shows the eigenvalues of the discriminant function as well as the canonical correlation for the discriminant function. The larger the eigenvalue, the more amount of variance shared in the linear combination of variables. Since only one function is involved, the function then explains majority of variance in the relationship. An eigenvalue of 0.414 and the percentage variance of 100 percent for the function indicates that, the derived discriminant function explains 100 percent variation in the relationship. This therefore reveals the importance of the discriminant function in the provision of maximum separation between the groups. Also since only one discriminant function was involved, the cumulative percentage of the variance was recorded as 100 percent. From the same table the Canonical correlation value was observed to be 0.541 and it explains an above average relationship between the discriminant scores and the levels of the dependent variable.

Wilks lambda

Wilks’ Lambda is the ratio of within-groups sums of squares to the total sums of squares. This is the proportion of the total variance in the discriminant scores not explained by differences among groups. A small lambda indicates that group means appear to differ. The Wilks lambda value of 0.707 from Table 5 indicates that not all of the independent variables contribute significance in the function. The table also provides a Chi-Square statistic to test the significance of Wilk’s Lambda.

As evident from the table, the Wilks lambda of 0.707, the chi-square statistics of 212.812 with $p$-value of 0.000 is less than the significant level ($\alpha$) of 0.05 and hence, the derived discriminant function explains the group membership well, thus, the group means appear to differ.

Table 6 summarises the output of the standardised canonical discriminant function coefficient and the structure matrix. The standardised canonical discriminant function coefficient was used to rank the importance of each of the seven independent variables. From Table 6, the standardised canonical discriminant function coefficient for the age and BMI variables was observed to be 0.979 and 0.307 respectively.
This means that, the group separation depends mostly on the age and BMI of the patients. In other words a patient is being diagnosed as being type 1 or type 2 diabetic status based on their age and measured BMI. Other variables that the group separation depended on were the patient’s height, systolic Blood pressure and the diastolic blood pressure. From Table 6, the canonical structure matrix revealed the correlations between each variable in the model and the discriminant function. It is expected that, a variable with correlation of 0.3 or more is considered to be very important. Similarly the age of patients was observed to be a major determining factor in classifying a patient as either type 1 or type 2 since a strong correlation of 0.989 between the ages and the function was observed. Also a weak positive correlation between the BPS, Ht, DBP, Wt, BMI and the function was observed.

The canonical discriminant function coefficients/ Fishers LinearDiscriminant Function obtained from the study was:

\[ D_1 = -7.176 + 0.078 \text{Age} - 0.024 \text{Wt} + 1.746 \text{Ht} + 0.004 \text{BPS} - 0.001 \text{BPD} - 0.025 \text{FBS} + 0.051 \text{BMI} \] (11)

Hence based on the derived Fishers Discriminant Function, the classification rule for the two diabetes patient groups were obtained and was used to compute the discriminant scores for classifying the original observations into their respective groups. The classification rules obtained are:

\[ D_1 = -112.814 + 0.25 \text{Age} - 12.39 \text{Wt} + 135 \text{Ht} + 0.384 \text{BPS} + 0.08 \text{DBP} + 0.24 \text{FBS} + 32.18 \text{BMI} \] (12)
\[ D_2 = -114.191 + 0.42 \text{Age} - 124.5 \text{Wt} + 135535 \text{Ht} + 0.39 \text{BPS} + 0.08 \text{DBP} + 0.18 \text{FBS} + 32.29 \text{BMI} \] (13)

From Table 7, the computed discriminant scores for Type 1 \((D_1)\) and Type 2 \((D_2)\) diabetes groups were able to correctly classify 85.3 percent of the original observations into their respective groups. Also based on the error rates obtained by the cross-validation method as evident in Table 7, 84.8 percent of the cross validated grouped cases were correctly classified. The results shows a clear indication that, the derived FLDF as well as the classification rules provided maximum separation between the two main diabetes group patients (i.e. either a type 1 or type 2 diabetic patient).

4.0 Conclusion

This study focused on deriving a discriminant function based on some identified variables in providing maximum separation between two groups of diabetes patients at Komfo Anokye teaching Hospital. Fishers Linear Discriminant Function based on the seven measured variables as well as the corresponding classification rule were developed. From the study 85.3 percent of the original observations were correctly classified whilst 84.8 percent of cross-validated observation were correctly classified. Also the classification of the patients into their respective diagnosed diabetes status depended hugely on the patient’s age as well as their BMI and to some small extent their FBS. The derived linear discriminant function provided maximum separation and the classification rule obtained will be use to classify future diabetic patients with similar identified variables as whether the person will belong to a type 1 or type 2 diabetes status group.
5.0 References


6.0 Tables

Table 1: Box’s M Test of Equality of Covariance Matrices

<table>
<thead>
<tr>
<th>Groups (DM Types)</th>
<th>Rank</th>
<th>Log determinants</th>
<th>Test Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type 2</td>
<td>7</td>
<td>22.107</td>
<td>F Approx.</td>
</tr>
<tr>
<td>Type 1</td>
<td>7</td>
<td>21.971</td>
<td>df1</td>
</tr>
<tr>
<td>Pooled within</td>
<td>7</td>
<td>22.553</td>
<td>df2</td>
</tr>
</tbody>
</table>

Table 2: TestorEqualityofMeanVectors

<table>
<thead>
<tr>
<th>Hotelling$T^2$</th>
<th>Test Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>255.7643</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Pooled within Group Matrices

<table>
<thead>
<tr>
<th>Age</th>
<th>Wt</th>
<th>Ht</th>
<th>BPS</th>
<th>DBP</th>
<th>FBS</th>
<th>BMI</th>
</tr>
</thead>
<tbody>
<tr>
<td>159.185</td>
<td>11.193</td>
<td>.56</td>
<td>4.026</td>
<td>-.220</td>
<td>-6.350</td>
<td>1.656</td>
</tr>
<tr>
<td>11.193</td>
<td>238.951</td>
<td>.421</td>
<td>52.837</td>
<td>36.374</td>
<td>12.026</td>
<td>75.701</td>
</tr>
<tr>
<td>.056</td>
<td>.421</td>
<td>.009</td>
<td>.014</td>
<td>-.100</td>
<td>-.008</td>
<td>-.162</td>
</tr>
<tr>
<td>4.026</td>
<td>52.837</td>
<td>.014</td>
<td>518.792</td>
<td>164.189</td>
<td>12.026</td>
<td>17.114</td>
</tr>
<tr>
<td>-.220</td>
<td>36.374</td>
<td>-.100</td>
<td>164.189</td>
<td>1009.707</td>
<td>-.818</td>
<td>16.262</td>
</tr>
<tr>
<td>-.6350</td>
<td>-4.579</td>
<td>-.008</td>
<td>12.026</td>
<td>-.818</td>
<td>23.287</td>
<td>-1.350</td>
</tr>
<tr>
<td>1.656</td>
<td>75.701</td>
<td>-.162</td>
<td>17.114</td>
<td>16.262</td>
<td>-1.350</td>
<td>36.499</td>
</tr>
</tbody>
</table>

Table 4: EigenValues

<table>
<thead>
<tr>
<th>Function(s)</th>
<th>Eigenvalue</th>
<th>% of Variance</th>
<th>Cumulative %</th>
<th>Canonical correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.414</td>
<td>100.0</td>
<td>100.0</td>
<td>0.541</td>
</tr>
</tbody>
</table>
Table 5: Wilk’s Lambda

<table>
<thead>
<tr>
<th>Test of Function(s)</th>
<th>Wilks' Lambda</th>
<th>Chi-square</th>
<th>df</th>
<th>P-value</th>
<th>Significance level(α)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.707</td>
<td>212.812</td>
<td>7</td>
<td>.000</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Table 6: Table of Standardized Canonical Discriminant Function Coefficients and Structure Matrix

Standardise canonical discriminant function coefficient

<table>
<thead>
<tr>
<th>Variables</th>
<th>Function</th>
<th>Variables</th>
<th>Structure Matrix</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>.979</td>
<td>Age</td>
<td>.986</td>
<td></td>
</tr>
<tr>
<td>Weight</td>
<td>-.364</td>
<td>FBS</td>
<td>-.206</td>
<td></td>
</tr>
<tr>
<td>Height</td>
<td>.165</td>
<td>SBP</td>
<td>.088</td>
<td></td>
</tr>
<tr>
<td>SBP</td>
<td>.102</td>
<td>Height</td>
<td>.022</td>
<td></td>
</tr>
<tr>
<td>DBP</td>
<td>.004</td>
<td>DBP</td>
<td>.021</td>
<td></td>
</tr>
<tr>
<td>FBS</td>
<td>-.121</td>
<td>Weight</td>
<td>.012</td>
<td></td>
</tr>
<tr>
<td>BMI</td>
<td>.307</td>
<td>BMI</td>
<td>.005</td>
<td></td>
</tr>
</tbody>
</table>

Table 7: Classification Results

<table>
<thead>
<tr>
<th>DM TYPES</th>
<th>Predicted Group Membership</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Type 1</td>
<td>Type 2</td>
</tr>
<tr>
<td>Original</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Count</td>
<td></td>
<td></td>
</tr>
<tr>
<td>%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type 1</td>
<td>57</td>
<td>2</td>
</tr>
<tr>
<td>Type 2</td>
<td>89</td>
<td>472</td>
</tr>
<tr>
<td>Cross-validated</td>
<td>Type 1</td>
<td>Type 2</td>
</tr>
<tr>
<td>Count</td>
<td></td>
<td></td>
</tr>
<tr>
<td>%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type 1</td>
<td>57</td>
<td>2</td>
</tr>
<tr>
<td>Type 2</td>
<td>92</td>
<td>469</td>
</tr>
<tr>
<td>Type 2</td>
<td>16.4</td>
<td>83.6</td>
</tr>
</tbody>
</table>

DM = Diabetes Mellitus