Jurnal Matematika dan Statistika serta Aplikasinya, Vol. 13, No. 2, 1-12, 2025
DOI:https://doi.org/10.24252/msa.v13i2.59040

Exploring Diabetes Mellitus Risk Patterns with Multiple Correspondence Analysis at Torabelo Hospital, Central Sulawesi

Asfar[1]*, Fadjryani[2], Valina Ambarwati B.Sarabi[2], Miftahul Jannah[2], Dewi Sartika[2], Aldi Setiawan[2], Khalilah Nurfadilah[3]

[1] *Data Science Study Program, Universitas Tadulako, Indonesia*
[2] *Statistics Study Program, Universitas Tadulako, Indonesia*
[3] *Mathematics Study Program, Universitas Islam Negeri Alauddin Makassar, Indonesia*
***Corresponding author**: asfar@untad.ac.id

***Submission date**: 4 July 2025, **Revision**: 3 August 2025, **Accepted**: 3 September 2025

ABSTRACT
This study aims to explore the pattern of diabetes mellitus risk factors among patients at Torabelo Regional Hospital in Sigi using Multiple Correspondence Analysis (MCA). A total of 465 patients were analyzed based on eight categorical variables, including age, gender, blood glucose levels, and lipid profiles. MCA was applied to identify inter-category relationships and visualize them in a low-dimensional space. The results show that most diabetes patients were female, aged 46 years and above, and had high fasting glucose, low HDL, and high LDL levels. The analysis identified two main patterns: a group with a low-risk metabolic profile who were not diagnosed with diabetes, and a group with a combination of high-risk metabolic categories who were more likely to already have diabetes. A distinct subgroup with extremely high triglyceride levels was also identified, indicating a rare but significant metabolic pattern. The first two dimensions of the MCA explained more than 40% of the data variation, providing sufficient support for meaningful visual interpretation. These findings demonstrate that MCA is effective in simplifying complex categorical data and supports risk-based segmentation strategies for early intervention planning in primary healthcare services, particularly in regions with high diabetes prevalence.

KEYWORDS
Diabetes Mellitus, Metabolic Risk Factors, Multiple Correspondence Analysis.

1. INTRODUCTION

Diabetes mellitus is a chronic high blood sugar condition that leads to various complications involving the eyes, kidneys, nerves, and blood vessels [1]. Insulin plays a crucial role in helping body cells utilize glucose as an energy source. In individuals with diabetes mellitus, impaired insulin production or function results in inefficient glucose utilization, causing it to accumulate in the bloodstream. This condition is known as hyperglycemia and is a hallmark of diabetes, particularly type 2. If left untreated, diabetes mellitus can lead to chronic hyperglycemia, which may result in organ damage and severe complications such as diabetic coma [2].

Sigi Regency in Central Sulawesi is experiencing a significant burden of diabetes. In 2021, there were 15,244 recorded cases of diabetes mellitus, yet only 704 patients were actively receiving treatment, indicating a substantial treatment gap. Furthermore, RSUD Torabelo reported an increase in type 2 diabetes mellitus cases from 466 in 2020 to 526 in 2021, suggesting a rising trend in disease prevalence in the region [3].

Risk factors for diabetes mellitus are categorized into non-modifiable factors such as age and gender, and modifiable factors

including blood glucose and lipid profile (HDL, LDL, and triglycerides) [4]. Prolonged high blood glucose levels can disrupt lipid metabolism, elevate triglyceride levels, and increase the risk of cardiovascular disease [5]. Lipid imbalances, especially when combined with hyperglycemia, may exacerbate diabetes mellitus complications. Therefore, it is essential to examine how these clinical risk factors are interrelated in diabetic patients.

Multiple Correspondence Analysis (MCA) is an exploratory multivariate method suitable for categorical data. Unlike principal component analysis (PCA) or factor analysis, which focus on numerical data, MCA maps categorical variables into two or three-dimensional space, allowing the relationships between categories to be visualized based on their proximity [6]. Each category is represented as a point, and closely located points indicate similar distribution patterns. MCA is advantageous because it does not require strict assumptions such as normality or linearity [7]. Compared to the chi-square test, MCA is more efficient in handling interactions among several variables simultaneously [8].

Previous studies have applied MCA in diabetes-related research. One study explored lifestyle risk factors and national diabetes care systems in 28 European Union countries, although the focus was on cross-country comparisons [9]. Another study applied MCA to investigate risk factors for diabetic foot, but the analysis was limited to specific clinical aspects [10]. However, no study has comprehensively explored the relationships among multiple clinical categories in diabetic patients using MCA. Conventional analytical approaches are still commonly used in diabetes mellitus studies, such as logistic regression to identify factors like obesity, hypertension, and physical activity associated with diabetes mellitus [11]. In addition to logistic regression, the chi-square test has also been used in diabetes mellitus cases [12].

Based on this background, the objectives of this study are: to analyze the associations among categories of diabetes mellitus risk factors based on patient data from RSUD Torabelo Sigi; to identify categories of risk factors that tend to occur together in certain respondent groups; and to visualize the results in a low-dimensional correspondence map using MCA. These findings are expected to support targeted policy-making and early intervention planning at the primary health care level, particularly for diabetes prevention in Sigi Regency.

2. LITERATURE REVIEW

2.1 Correspondence Analysis

Correspondence analysis was first developed in the 1960s by a French statistician named Jean-Paul Benzecri. Benzecri introduced the concept of correspondence analysis as a method to analyze two-way or multi-way contingency tables and visualize the relationships between categorical variables. Correspondence analysis posits that graphical data representation offers several advantages, including data summarization, ease of interpretation due to simplified data presentation, and visual data display [13]. This method projects the rows and columns of a data matrix as points in a low-dimensional graph. Correspondence analysis is used to reduce the dimensionality of variables and depict the profiles of row and column vectors in a contingency table data matrix. The results of correspondence analysis typically utilize the two best dimensions to represent the data, which serve as point coordinates and a measure of the amount of information contained in each dimension, commonly referred to as inertia.

Correspondence analysis is a non-parametric technique, meaning it does not require assumption testing, such as normality, autocorrelation, multicollinearity, heteroscedasticity, or linearity. The variables used in correspondence analysis are discrete, consisting of nominal or ordinal data with multiple categories.

Correspondence analysis has several advantages, including its suitability for analyzing categorical variable data that can be simply represented through cross-tabulation, its ability to depict relationships not only between rows and columns but also between categories within each row and column, its provision of a combined graphical display of row and column categories in a single figure with the same dimensions, its applicability to non-linear relationships, and its flexibility for use with large data matrices. Despite its advantages, correspondence analysis also has limitations. These include its unsuitability for hypothesis testing, though it is highly appropriate for data exploration, and the lack of a specific method to determine or decide the optimal number of dimensions [13].

2.2 Multiple Correspondence Analysis

Multiple Correspondence Analysis (MCA) is a multivariate exploratory data analysis technique for two or more variables projected simultaneously in a graph with rows and columns. MCA is an extension of Joint Correspondence Analysis (JCA) [13].

In JCA, the solution is considered to overestimate the total inertia value because the Burt matrix structure, with submatrices on its main diagonal, consists of cross-tabulations of each variable with itself. To address this issue, MCA emerged as an alternative. MCA is an iterative algorithm that identifies the optimal weighted least squares solution corresponding to the off-diagonal table of the Burt matrix, aiming to maximize the proportion of inertia value to achieve greater accuracy in data processing results.

2.3 Biplot Analysis

One of the multivariate methods widely used to facilitate the interpretation of large datasets with numerous variables is biplot analysis. Biplot analysis is a statistical method where the results are presented visually, aiming to simultaneously represent n observational objects and p variables in a two-dimensional plane. Additionally, biplot displays provide four key insights: the proximity between objects, the variability of variables, the relationships (correlations) between variables, and the values of variables for a specific object. Biplot analysis is based on the singular value decomposition of a matrix.

3. METHODOLOGY

3.1 Data and Variables

The data used in this study were obtained from medical records and laboratory test results of diabetes mellitus patients at Torabelo Hospital, Sigi Regency, Central Sulawesi, in 2022. The study utilized eight categorical variables: gender, age, diabetes mellitus status, fasting blood glucose, 2-hour postprandial blood glucose, High-Density Lipoprotein (HDL), Low-Density Lipoprotein (LDL), and triglycerides [14], [15]. The details of the variables are depicted in **Table 1**.

3.2 Analysis Procedure

The steps involved in the multiple correspondence analysis (MCA) conducted in this study are as follows:

1. Indicator Matrix

   The indicator matrix, commonly denoted as $\mathbf{Z}$ is a matrix that indicates the presence of each category for every individual or case. The indicator matrix $\mathbf{Z}$ contains binary elements, either 0 or 1. The matrix $\mathbf{Z}$ has dimensions $N \times J$ where N is the total number of individuals and J is the total number of categories [16].

2. Burt Matrix

   The Burt matrix is a multi-way contingency table resulting from the cross-tabulation of the combined indicator matrix of categorical variables. If there are W categorical variables, then $\mathbf{Z}_w$ is the indicator matrix for the w-th categorical variable, with $w = 1, 2, \ldots, W$ meanwhile $\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 & \mathbf{Z}_2 & \ldots & \mathbf{Z}_W \end{bmatrix}$ represents the combined indicator matrix of all categorical variables. The matrix $\mathbf{Z}$ is then cross-tabulated as $\mathbf{Z}^T\mathbf{Z}$ according to equation (1) [13].

$$\mathbf{B} = (b_{ij}) = \mathbf{Z}^\top \mathbf{Z} = \begin{bmatrix} \mathbf{Z_1}^\top \mathbf{Z_1} & \mathbf{Z_1}^\top \mathbf{Z_2} & \cdots & \mathbf{Z_1}^\top \mathbf{Z_W} \\ \mathbf{Z_2}^\top \mathbf{Z_1} & \mathbf{Z_2}^\top \mathbf{Z_2} & \cdots & \mathbf{Z_2}^\top \mathbf{Z_W} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{Z_W}^\top \mathbf{Z_1} & \mathbf{Z_W}^\top \mathbf{Z_2} & \cdots & \mathbf{Z_W}^\top \mathbf{Z_W} \end{bmatrix} \tag{1}$$

3. Correspondence Matrix

   The correspondence matrix is a matrix whose elements are derived from the elements of the Burt matrix, divided by n, where n is the total number of samples and W is the number of variables used.

$$\mathbf{P} = \frac{\mathbf{Z}}{n \times W} \tag{2}$$

4. Row Mass and Column Mass

   Row mass and column mass represent the proportion of a particular category relative to all formed categories. The row mass r_i is obtained by summing each row of the correspondence matrix, using the equation (3):

$$\mathbf{r_i} = \mathbf{P}.\mathbf{1}_{i \times 1} \tag{3}$$

Table 1. Research Variables

Variable	Operational Definition	Category
Diabetes Mellitus	High blood sugar levels accompanied by impaired metabolism of carbohydrates, fats, and proteins as a result of insulin function disorders.	0 : No Diabetes 1 : Diabetes
Gender	Gender assigned at birth.	1 : Men 2 : Female
Age	The number of years from birth to the most recent birthday.	1 : $<$ 26 Years 2 : 26 – 45 Years 3 : 46 – 65 Years Old 4 : $>$ 65 Years Old
2-Hour Postprandial Blood Glucose	Blood glucose measured 2 hours after eating.	1 : Normal ($<$ 140 mg/dL) 2 : Prediabetes (140 – 199 mg/dL) 3 : Diabetes ($\geq$ 200 mg/dL)
Fasting Blood Glucose	Blood glucose level measured after fasting.	1 : Normal ($<$ 100 mg/dL) 2 : Prediabetes (100 – 125 mg/dL) 3 : Diabetes ($\geq$126 mg/dL)
High-Density Lipoprotein (HDL)	High-density lipoprotein, mainly composed of proteins.	1 : Low ($<$ 40 mg/dL) 2 : Normal (40 – 59 mg/dL) 3 : High ($\geq$ 60 mg/dL)
Low-Density Lipoprotein (LDL)	Lipoprotein that carries the most cholesterol in the body.	1 : Low ($<$ 100) 2 : Medium (100 – 129) 3 : Height ($\geq$ 130)
Trigliserida	The body's energy reserves that can be derived from fatty foods and carbohydrates.	1 : Normal ($<$ 150 mg/dL) 2 : Moderate (150 – 199 mg/dL) 3 : High ($\geq$ 200 mg/dL)

The row diagonal matrix $\mathbf{D_r}$ is then derived as follows:

$$\mathbf{D_r} = diag\ (\mathbf{r}) = \begin{bmatrix} \mathbf{p_{1.}} & 0 & \dots & 0 \\ 0 & \mathbf{p_{2.}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{p_{i.}} \end{bmatrix} \tag{4}$$

The column mass $\mathbf{c}_j$ is calculated in the same way, by summing each column of the correspondence matrix.

$$\mathbf{c}_i = \mathbf{P}.\mathbf{1}_{i\times\mathbf{1}}^T \tag{5}$$

The column diagonal matrix $\mathbf{D}_c$ is obtained as follows:

$$\mathbf{D_c} = diag\ (\mathbf{c}) = \begin{bmatrix} \mathbf{p_{.1}} & 0 & \dots & 0 \\ 0 & \mathbf{p_{.2}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{p_{.j}} \end{bmatrix} \tag{6}$$

5. Row Profile and Column Profile

A profile represents the proportion of each row or column in the correspondence matrix, where each observed frequency in row i and column j is divided by the respective total of that row or column. The matrix $\mathbf{R}$ of size $a \times b$ can then be formed as follows:

$$\mathbf{R} = \mathbf{D}_r^{-1}\mathbf{P} = \begin{bmatrix} \frac{\mathbf{p}_{11}}{p_{1.}} & \frac{\mathbf{p}_{12}}{\mathbf{p}_{1.}} & \cdots & \frac{\mathbf{p}_{1b}}{\mathbf{p}_{1.}} \\ \frac{\mathbf{p}_{21}}{\mathbf{p}_{2.}} & \frac{\mathbf{p}_{22}}{\mathbf{p}_{2.}} & \cdots & \frac{\mathbf{p}_{2b}}{\mathbf{p}_{2.}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\mathbf{p}_{a1}}{\mathbf{p}_{a.}} & \frac{\mathbf{p}_{a2}}{\mathbf{p}_{a.}} & \cdots & \frac{\mathbf{p}_{ab}}{\mathbf{p}_{a.}} \end{bmatrix} \tag{7}$$

Matrix $\mathbf{R}$ is referred to as the row profile in a b-dimensional space, where the sum of elements in each row profile equals one. Meanwhile, the matrix $\mathbf{C}$ of size $b \times a$ is defined as:

$$\mathbf{C} = \mathbf{D}_c^{-1}\mathbf{P} = \begin{bmatrix} \frac{\mathbf{p}_{11}}{p_{1.}} & \frac{\mathbf{p}_{12}}{\mathbf{p}_{1.}} & \cdots & \frac{\mathbf{p}_{1b}}{\mathbf{p}_{1.}} \\ \frac{\mathbf{p}_{21}}{\mathbf{p}_{2.}} & \frac{\mathbf{p}_{22}}{\mathbf{p}_{2.}} & \cdots & \frac{\mathbf{p}_{2b}}{\mathbf{p}_{2.}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\mathbf{p}_{a1}}{\mathbf{p}_{b.}} & \frac{\mathbf{p}_{a2}}{\mathbf{p}_{b.}} & \cdots & \frac{\mathbf{p}_{ab}}{\mathbf{p}_{b.}} \end{bmatrix} \tag{8}$$

Matrix $\mathbf{C}$ is known as the column profile in an a-dimensional space, where the sum of elements in each column profile equals one.

6. Standardized Residual Matrix

   The standardized residual matrix ($\mathbf{S}$) is obtained using the following equation (9):

$$\mathbf{S} = \mathbf{D}_r^{-\frac{1}{2}} \left(\mathbf{P} - \mathbf{r}\mathbf{c}^{\top} \right) \mathbf{D}_c^{-\frac{1}{2}} \tag{9}$$

7. Singular Value Decomposition (SVD)

   The singular value decomposition of the standardized residual matrix ($\mathbf{S}$) can be performed using the following equation:

$$\mathbf{A} = \mathbf{U}\mathbf{D}_{\lambda}\mathbf{V}^T \tag{10}$$

   with $\mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = \mathbf{I}$ Where $\mathbf{U}$ is an orthogonal matrix of size $m \times m$, $\mathbf{D}_{\lambda}$ is a diagonal matrix of size $m \times m$ whose elements are the singular values of $\mathbf{S}$, and $\mathbf{V}$ is an orthogonal matrix of size $n \times n$. $\mathbf{I}$ represents the identity matrix. he result obtained from this decomposition consists of eigenvectors $\mathbf{v}_{ij}$ and singular values $\mathbf{a}_s$ or each formed dimension [13].

8. Principal Inertia Values The principal inertia values (λ_s) indicates the contribution of the s-th dimension to the total inertia. It is calculated using the following equation:

$$\lambda_s = \mathbf{a}_s{}^2 \tag{11}$$

   The total inertia (Λ) is then obtained by calculating the trace of the product of the standardized residual matrix and its transpose.

$$\Lambda = trace\ \left(\mathbf{S}^{\mathbf{T}}\mathbf{S}\right) = \sum_{i=1}^{I}\sum_{j=1}^{J} \frac{\left(P_{ij} - r_i c_j\right)^2}{r_i c_j} \tag{12}$$

9. Profile Coordinates

   In this analysis, profile coordinates are divided into two types: standard coordinates and principal coordinates.

a. Standard Coordinates

Standard coordinates are a set of point coordinates where the weighted sum of squares along each axis equals one [16]. The standard coordinates are calculated using the following equations:

$$\mathbf{X} = \mathbf{D}_r^{-\frac{1}{2}}\mathbf{U} \tag{13}$$

$$\mathbf{Y} = \mathbf{D}_c^{-\frac{1}{2}}\mathbf{V} \tag{14}$$

Where $\mathbf{X}$ represents the row standard coordinates and Y represents the column standard coordinates.

b. Principal Coordinates

Principal coordinates refer to the coordinates of a set of points projected onto the principal axes, such that the weighted sum of squares along each principal axis equals the principal inertia value for that axis. Row principal coordinates are used to describe the distribution of respondents, while column principal coordinates represent the categorical variables. The principal coordinates for rows ($\mathbf{F}$) and columns ($\mathbf{G}$) are obtained using the following equations [13]:

$$\mathbf{F} = \mathbf{D}_r^{-\frac{1}{2}}\mathbf{U}\boldsymbol{\Sigma} \tag{15}$$

$$\mathbf{G} = \mathbf{D}_c^{-\frac{1}{2}}\mathbf{V}\boldsymbol{\Sigma} \tag{16}$$

4. RESULT & DISCUSSION

4.1 Descriptive Statistics

Based on **Table 2**, it was found that the majority of patients at Torabelo Regional General Hospital in Sigi fall into the high metabolic risk group. This finding is consistent with previous studies, which reported that most diabetic patients are aged 45 years or older [17]. Physiologically, the risk of diabetes increases with age due to reduced muscle mass, increased body fat, and decreased pancreatic β-cell function, which affects insulin secretion [18].

In addition, women have a slightly higher proportion of diabetes cases than men. This is consistent with previous findings that explain women, especially postmenopausal women, are more susceptible to diabetes due to hormonal changes and increased body mass index [19]. A similar pattern was also observed in a more recent local study that reported a higher prevalence of diabetes among female respondents [17].

Most patients had 2-hour blood glucose levels in the prediabetes category (42.8%), while as many as 38.3% of patients had already entered the diabetes category based on fasting blood glucose levels. This condition indicates that many patients are already in a serious metabolic disorder phase. A person is considered prediabetic when their blood glucose levels exceed the normal range but have not yet reached the diabetes threshold [20]. Meanwhile, fasting blood glucose levels equal to or greater than 126 mg/dL are typically classified as hyperglycemia or diabetes mellitus [21].

4.2 Eigenvalues and Proportion of Variance

Multiple correspondence analysis was used to examine the relationships between categories of several variables simultaneously. In this study, the number of dimensions to be used in depicting the perceptual map will be determined. The **Table 3** depict the results of the analysis to determine the appropriate dimensional size for plotting the multiple correspondence analysis

Based on **Table 3**, it was found that the first two dimensions account for more than 40% of the total inertia. This indicates that using two dimensions is sufficient to represent the underlying structure of the data. This is in line with the general recommendation in multiple correspondence analysis, which states that the first two or three dimensions are usually adequate for visual exploration and initial interpretation, especially when the proportion of explained inertia exceeds 40% [16]. If necessary,

Table 2. Descriptive Statistics

Variable	Category	Percentage
Diabetes Mellitus	0 : No Diabetes	33.8%
	1 : Diabetes	66.2%
Gender	1 : Men	47.3%
	2 : Female	52.7%
Age	1 : $<$ 26 Years	4.5%
	2 : 26 – 45 Years	20.2%
	3 : 46 – 65 Years Old	64.5%
	4 : $>$ 65 years old	10.8%
2-Hour Postprandial Blood Glucose	1 : Normal ($<$ 140 mg/Dl)	39.6%
	2 : Prediabetes (140 – 199 mg/dL)	42.8%
	3 : Diabetes ($\geq$ 200 mg/dL)	17.6%
Fasting Blood Glucose	1 : Normal ($<$ 100)	33.3%
	2 : Prediabetes (100 – 125 mg/dL)	28.4%
	3 : Diabetes ($\geq$ 126 mg/dL)	38.3%
High-Density Lipoprotein (HDL)	1 : Low ($<$ 40 mg/dL)	39.6%
	2 : Normal (40 – 59 mg/dL)	36.8%
	3 : High ($\geq$ 60 mg/dL)	23.7%
Low-Density Lipoprotein (LDL)	1 : Low ($<$ 100)	32.5%
	2 : Medium (100 – 129)	27.1%
	3 : Height ($\geq$ 130)	40.4%
Trigliserida	1 : Normal ($<$ 150 mg/dL) 2 : Moderate (150 – 199 mg/dL) 3 : High ($\geq$ 200 mg/dL)	43.2% 48.2% 8.6%

Table 3. Eigenvalues and Proportion of Variance

Dimensions	Eigenvalue	Variance Percent	Cumulative Variance Percent
Dim.1	0.5706	30.4300	30.4301
Dim.2	0.2251	12.0076	42.4376
Dim.3	0.1502	8.0125	50.4501
Dim.4	0.1276	6.8032	57.2532
Dim.5	0.1267	6.7558	64.0090
Dim.6	0.1246	6.6464	70.6554
Dim.7	0.1104	5.8872	76.5427
Dim.8	0.1026	5.4742	82.0169
Dim.9	0.0900	4.8018	86.8187
Dim.10	0.0704	3.7564	90.5751
Dim.11	0.0630	3.3606	93.9357
Dim.12	0.0446	2.3796	96.3154
Dim.13	0.0348	1.8578	98.1732
Dim.14	0.0217	1.1569	99.3300
Dim.15	0.0126	0.6699	100.0000

researchers may use the first three dimensions to perform three-dimensional visualization or further multivariate analysis. However, two dimensions are generally sufficient to identify patterns and map relationships between categories. This reflects the fundamental purpose of MCA, which is to simplify complex categorical data into a visual structure that can be understood without losing essential information [22].

4.3 Profile Coordinates

Next, profile coordinates will be constructed, in this case the principal column coordinates, to create the perceptual map. The **Table 4** depict the results of the analysis determining the principal column coordinates for each level of categorical variables used in this study.

Table 4. Principal Column Coordinates

Category	Dimension 1	Dimension 2	Dimension 3
Diabetes Mellitus	0.012	-0.123	-0.506
Gender	-0.011	0.111	0.454
Age	0.885	0.346	3.058
2-Hour Postprandial Blood Glucose	0.943	-0.117	-0.395
Fasting Blood Glucose	-0.278	0.017	-0.028
High-Density Lipoprotein (HDL)	-0.478	-0.026	-0.377
Low-Density Lipoprotein (LDL)	0.913	0.010	0.215
Trigliserida	-0.696	-0.296	0.306
Diabetes Mellitus	-0.361	0.696	-1.225
Gender	1.307	-0.035	-0.059
Age	-0.691	-0.714	0.466
2-Hour Postprandial Blood Glucose	-0.626	0.559	-0.294
Fasting Blood Glucose	-0.629	0.649	0.288
High-Density Lipoprotein (HDL)	1.122	0.033	-0.156
Low-Density Lipoprotein (LDL)	-0.691	-1.137	-0.240
Trigliserida	1.337	-0.046	-0.054
Diabetes Mellitus	-0.692	-0.983	-0.063
LDL (2)	-0.610	0.695	0.085
Trigliserida (0)	0.881	-0.039	-0.070
Trigliserida (1)	-0.696	-0.348	0.004
Trigliserida (2)	-0.525	2.146	0.329
Diabetes (0)	1.151	-0.107	0.151
Diabetes (1)	-0.587	0.055	-0.077

Based on **Table 4**, the principal column coordinates derived from the multiple correspondence analysis (MCA) can be used to observe which categories contribute to the formation of the main dimensions in the data. In Dimension 1, several categories show relatively high coordinate values, including age under 26 years at 0.88490, age between 26–45 years at 0.9429, 2-hour blood glucose in the normal category at 0.9128, fasting blood glucose in the normal category at 1.3068, High-Density Lipoprotein (HDL) in the normal category at 1.1217, Low-Density Lipoprotein (LDL) in the low category at 1.3369, triglycerides in the normal category at 0.8805, and not having diabetes mellitus at 1.1509. These categories contribute strongly and positively to the first dimension. Conversely, categories such as fasting blood glucose in the prediabetes category and triglycerides in the moderate category have relatively high negative values, at -0.6909 and -0.6963 respectively. This indicates that these variables are located at the opposite end of the spectrum compared to the positively associated categories.

From a clinical interpretation, Dimension 1 likely represents the contrast between patients with low-risk and high-risk metabolic profiles. On the positive side are variables such as HDL in the normal category and LDL in the low category, which indicate more favorable lipid levels, as well as not having diabetes, indicating the absence of the disease. This aligns with

established clinical guidelines that define fasting blood glucose below 100 mg/dL and HDL levels between 40–59 mg/dL as within the normal and safe range [20]. Thus, the positive side of Dimension 1 can be interpreted as representing characteristics of patients with a lower risk of diabetes mellitus.

Categories with negative contributions to Dimension 1, such as fasting blood glucose in the prediabetes category and HDL in the high category, suggest the presence of early metabolic disturbances. Elevated fasting glucose levels indicate potential insulin resistance. While high HDL levels are conventionally considered beneficial, the placement of high HDL far from the reference cluster suggests that high HDL does not universally protect against diabetes mellitus. This is consistent with findings that certain HDL subfractions, particularly HDL-2 and HDL-L, tend to decrease in individuals with insulin resistance, implying that high total HDL levels do not always provide metabolic protection [23].

In Dimension 2, the most prominent contributor is triglycerides in the high category, with the highest coordinate value of 2.1461, followed by 2-hour blood glucose in the diabetes category, LDL in the high category, and HDL in the low category. These variables are indicators of advanced metabolic dysfunction, including hyperglycemia, dyslipidemia, and low HDL levels. Together, they represent a group of individuals at high risk for diabetes mellitus. This finding supports previous research indicating that significantly elevated triglyceride levels play a key role in the onset of insulin resistance and impaired glucose metabolism, where increases in both triglycerides and glucose mutually exacerbate metabolic disturbances [24].

4.4 Multiple Correspondence Plot

After obtaining the principal column coordinates, the next step is to construct a perceptual map of all categories across two dimensions.

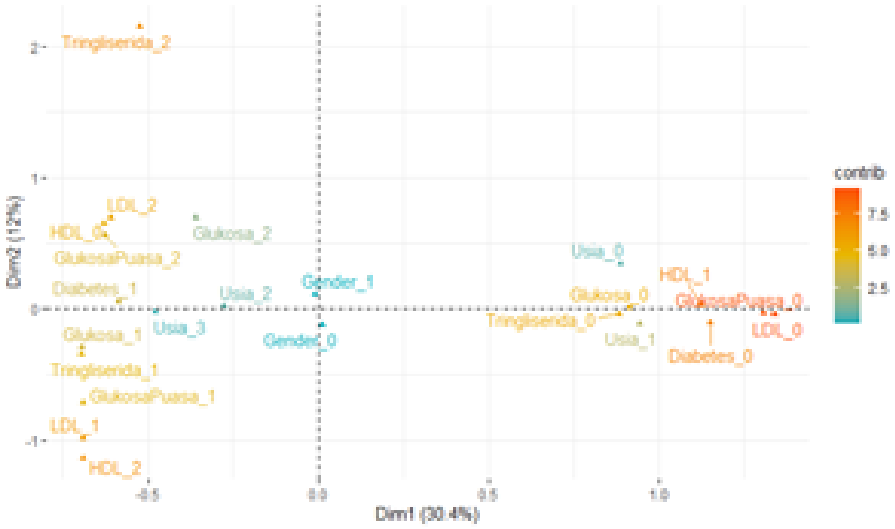

Figure 1. Perceptual Map of Variable Categories

Figure 1 presents the mapping of variable categories in a two-dimensional space. Dimension 1 explains 30.4% of the total inertia and captures the main pattern of relationships among the categories, while Dimension 2 contributes 12% of the inertia and provides additional information orthogonal to Dimension 1. Dimension 1 ranges from -0.5 to 1, while Dimension 2 ranges from -1 to 2. This scale indicates a fairly wide data distribution, with some categories located at extreme ends. The colored dots represent variable categories and their contribution to inertia, ranging from 0% (blue) to 7.5% (red). Categories in red (high contribution, $>5\%$), such as fasting blood glucose in the normal category and Low-Density Lipoprotein (LDL) in the low category, have significant influence, while categories in blue (low contribution, $<2.5\%$), such as age under 26 and gender (both male and female), contribute less.

In **Figure 1**, categories are grouped based on Dimension 1, into positive and negative sides. The positive side (right) includes categories such as age under 26, age between 26–45 years, triglycerides in the normal category, 2-hour blood glucose in the normal category, fasting blood glucose in the normal category, HDL in the normal category, LDL in the low category, and not having diabetes mellitus. These categories are positioned on Dimension 1 with relatively high positive values. This can be

seen from the similar colors of the dots representing each category, indicating co-occurrence patterns. These categories tend to appear among individuals who do not have diabetes mellitus. In addition, their position reflects that they are more likely to have glucose and lipid profiles within normal ranges and be of younger age.

Conversely, on the negative side of Dimension 1 (left), categories such as HDL in the low and high categories, fasting blood glucose in the diabetes category, 2-hour blood glucose in the diabetes category, triglycerides in the moderate category, age between 46–65 years, and age over 65 are clustered near the category of having diabetes mellitus. This indicates that these categories often appear together in individuals who are diagnosed with diabetes. This pattern suggests that they experience more metabolic disturbances, such as hyperglycemia and dyslipidemia, and belong to an older age group—all of which are major risk factors for diabetes.

Interestingly, the high triglycerides category appears as a separate point, positioned far from other categories, especially along Dimension 2, where it has a very high value. This indicates that the high triglycerides category does not frequently co-occur with other categories and may represent a unique subgroup within the population. Individuals in this subgroup are likely to have specific metabolic abnormalities. This interpretation is supported by previous research, which has shown that elevated triglyceride levels are strongly associated with insulin resistance and impaired glucose metabolism [24].

4.5 Multiple Correspondence Biplot

To provide deeper insight, the analysis continues with a Multiple Correspondence Analysis (MCA) biplot. The MCA biplot illustrates the relationship between variable categories and respondents, enabling a more detailed visual interpretation of groupings and profiles within the data. **Figure 2** displays a large number of dots representing individuals, grouped into two main

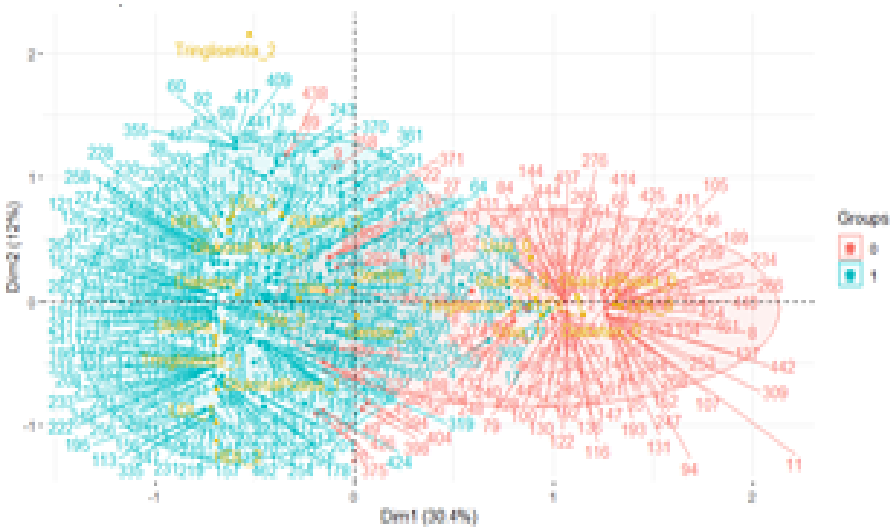

Figure 2. Multiple Correspondence Biplot

clusters: group 0 in red, indicating individuals who do not have diabetes mellitus, and group 1 in blue, representing individuals who have diabetes mellitus. The non-diabetic group is dominated by individuals located along Dimension 1 and tends to cluster together. In contrast, the diabetic group is more widely dispersed, with a distribution spanning from negative to positive values along Dimension 1 and ranging from -1 to 2 along Dimension 2. This pattern suggests greater heterogeneity within the diabetic group compared to the non-diabetic group. These results support the idea that the population can be segmented based on combinations of metabolic indicators, as outlined in the metabolic segmentation framework [25]. The non-diabetic group on the right side of the plot reflects a low-risk metabolic profile, whereas the left side indicates a high-risk profile. Meanwhile, the upper area of the plot reveals more unique and less common metabolic patterns.

Overall, the visual results support the conceptual framework of metabolic syndrome, which encompasses hyperglycemia, dyslipidemia (particularly low HDL and high LDL), and age-related risk factors. These conditions are recognized as a cluster of significant factors contributing to the onset of diabetes mellitus. This finding aligns with the theory of metabolic syndrome described in previous literature, which defines it as a constellation of metabolic abnormalities including high glucose, low HDL,

high triglycerides, high LDL, and advanced age interconnected through mechanisms of insulin resistancen [26]. Therefore, the category groupings in the MCA plot visually reflect the clinical structure of metabolic syndrome. Individuals with such combinations of variables tend to fall within the metabolically disturbed profile cluster that is susceptible to diabetes mellitus, whereas individuals with a normal metabolic profile are more frequently associated with non-diabetic status. Additionally, the unique position of the high triglycerides category, separated from the main cluster, may reflect a more extreme or rare form of hypertriglyceridemia but still relevant as a high-risk metabolic indicator.

From a methodological perspective, these results demonstrate the power of MCA as an exploratory technique that not only reduces the dimensionality of categorical data but also visualizes structural relationships among categories in an interpretable graphic form. As explained in previous literature, proximity between categories in the MCA space reflects similar distribution patterns in the population, whereas categories positioned far apart indicate distinct or rare occurrences [22].

5. CONCLUSION

This study shows that the majority of patients at Torabelo Regional General Hospital in Sigi are at high risk for diabetes mellitus, particularly those aged over 46 years and characterized by metabolic profiles such as elevated blood glucose, low High-Density Lipoprotein (HDL), and high Low-Density Lipoprotein (LDL). Multiple correspondence analysis successfully identified two main groups of respondents: one with a low-risk metabolic profile and not suffering from diabetes, and another group of individuals with a combination of high-risk metabolic categories who tend to already have diabetes, as well as a unique category with extremely high triglyceride levels indicating rare but extreme risk. These results logically support the study's claims by showing a consistent relationship between risk factors and diabetes status and clarifying the data structure through two-dimensional visualization. The development prospects of these findings include the implementation of metabolic profile-based screening models in primary healthcare services and the advancement of follow-up studies to evaluate specific interventions tailored to the identified risk clusters.

REFERENCES

[1] Indriyani, Ludiana, and T. K. Dewi, "Penerapan senam kaki diabetes melitus terhadap kadar glukosa darah pada penderita diabetes melitus di puskesmas yosomulyo," *Jurnal Cendikia Muda*, vol. 3, pp. 252–259,

[2] U. Hasanah, "Insulin sebagai pengatur kadar gula darah, hal," *Jurnal Keluarga Sehat Sejahterah*, vol. 11,

[3] D. K. S. Tengah, "Profil kesehatan provinsi sulawesi tengah tahun 2021," Dinas Kesehatan Sulawesi Tengah, Tech. Rep., 2023. [Online]. Available: https://dinkes.sultengprov.go.id/wp-content/uploads/2022/05/PROFIL-DINAS-KESEHATAN-2021.pdf

[4] R. Rosita, D. A. Kusumaningtiar, A. Irfandi, and I. M. Ayu, "Hubungan antara Jenis Kelamin, Umur, dan Aktifitas Fisik Lansia dengan Diabetes Melitus Tipe 2 di Puskesmas Balaraja Kabupaten Tangerang," *Jurnal Kesehatan Masyarakat (Undip)*, vol. 10, no. 3, pp. 364–371, May 2022.

[5] "Diabetes melitus: Review etiologi, patofisiologi, gejala, penyebab, cara pemeriksaan, cara pengobatan dan cara pencegahan," *UIN Alauddin Makassar*, vol. 1,

[6] M. Greenacre and T. Hastie, "The Geometric Interpretation of Correspondence Analysis," *Journal of the American Statistical Association*, vol. 82, no. 398, pp. 437–447, Jun. 1987.

[7] B. G. Tabachnick, L. S. Fidell, and J. B. Ullman, "Using multivariate statistics (7th ed.)," *Boston, MA: Pearson*, vol. 7th editio,

[8] "A correspondence analysis revealed frailty deficits aggregate and are multidimensional," *Journal of Clinical Epidemiology*, vol. 63,

[9] E. Altobelli, P. M. Angeletti, V. F. Profeta, and R. Petrocelli, "Lifestyle risk factors for type 2 diabetes mellitus and national diabetes care systems in european countries," *Nutrients*, vol. 12,

[10] "An exploration of diabetic foot screening procedures data by a multiple correspondence analysis," *Zdravstveno Varstvo*, vol. 56,

[11] N. K. Hasibuan, S. Dur, and I. Husein, "Faktor penyebab penyakit diabetes melitus dengan metode regresi logistik," *G-Tech: Jurnal Teknologi Terapan*, vol. 6,

[12] F. Nasution, A. Andilala, and A. A. Siregar, "Faktor Risiko Kejadian Diabetes Mellitus," *jik*, vol. 9, no. 2, p. 94,

[13] M. Greenacre, *Correspondence Analysis in Practice*, 0th ed. Chapman and Hall/CRC, May 2007. [Online]. Available: https://www.taylorfrancis.com/books/9781420011234

[14] American Diabetes Association, "*Standards of Medical Care in Diabetes—2022* Abridged for Primary Care Providers," *Clinical Diabetes*, vol. 40, no. 1, pp. 10–38, Jan. 2022.

[15] N. I. of Health, "ATP III Guidelines At-A-Glance Quick Desk Reference," U.S. Department of Health and Human Services, Tech. Rep., May 2001. [Online]. Available: https://www.nhlbi.nih.gov/resources/atp-iii-glance-quick-desk-reference

[16] A. Di Ciaccio, M. Coli, and J. M. Angulo Ibanez, Eds., *Advanced Statistical Methods for the Analysis of Large Data-Sets*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. [Online]. Available: http://link.springer.com/10.1007/978-3-642-21037-2

[17] V. R. Rohmatulloh, R. Riskiyah, B. Pardjianto, and L. S. Kinasih, "Hubungan Usia dan Jenis Kelamin terhadap Angka Kejadian Diabetes Mekitus Tipe 2 Berdasarkan 4 Kriteria Diagnosis di Poliklinik Penyakit Dalam RSUD Karsa Husada Kota Batu," *PREPOTIF : JURNAL KESEHATAN MASYARAKAT*, vol. 8, no. 1, pp. 2528–2543, Apr. 2024.

[18] Perkumpulan Endokrinologi Indonesia (PERKENI), *Pedoman Pengelolaan dan Pencegahan Diabetes Melitus Tipe 2 Dewasa di Indonesia*, S. A. Soelistijo, K. Suastika, D. Lindarto, E. Decroli, H. Permana, K. W. Sucipto, Y. Kusnadi, Budiman, M. R. Ikhsan, L. Sasiarini, H. Sanusi, K. H. N. HS, and H. Susanto, Eds. Jakarta: PB PERKENI, 2021, tim Penyusun Revisi 2021.

[19] K. Komariah and S. Rahayu, "Hubungan usia, jenis kelamin, dan indeks massa tubuh dengan kadar gula darah puasa pada pasien diabetes melitus tipe 2 di klinik pratama rawat jalan proklamasi, depok, jawa barat," *Jurnal Kesehatan Kusuma Husada*, vol. 11, no. 1, pp. 41–50,

[20] T. A. S. Ratih Puspita Febrinasari and S. E. P. Dyonisa Nasirochmi Pakha, *Buku Saku Diabetes Melitus untuk Awam*, 1st ed. Surakarta: UNS Press, 2020. [Online]. Available: https://www.researchgate.net/publication/346495581_BUKU_SAKU_DIABETES_MELITUS_UNTUK_AWAM

[21] World Health Organization (WHO), *Classification of Diabetes Mellitus*, 1st ed. Geneva: World Health Organization, 2019. [Online]. Available: https://www.who.int/publications/i/item/classification-of-diabetes-mellitus

[22] F. Husson, S. Lê, and J. Pagès, *Exploratory Multivariate Analysis by Example Using R*. Boca Raton, FL, United States: CRC Press, Taylor & Francis Group, 2011. [Online]. Available: https://www.taylorfrancis.com/books/mono/10.1201/b10345/exploratory-multivariate-analysis-example-using-francois-husson-sebastien-le-jrme-pags

[23] P. Piko *et al.*, "Association of hdl subfraction profile with the progression of insulin resistance," *International Journal of Molecular Sciences*, vol. 24, no. 17, p. 13563,

[24] W. R. Gumilar, "Hasil pemeriksaan kadar trigliserida dan kolesterol pada penderita diabetes melitus tipe 2 di rumah sakit efarina," *ULIL ALBAB: Jurnal Ilmiah Multidisiplin*, vol. 1, no. 5, pp. 1031–1038, 2022.

[25] D. Vistisen, D. R. Witte, E. J. Brunner, M. Kivimäki, A. Tabák, and M. E. Jørgensen, "Risk of cardiovascular disease and death in individuals with prediabetes defined by different criteria: The whitehall ii study," *Diabetes Care*, vol. 41, no. 4, pp. 899–906,

[26] V. Blaton, "How is the metabolic syndrome related to the dyslipidemia?" *EJIFCC*, vol. 18, no. 1, pp. 15–22, 2007.