DOI 10.26773/smj.240210

ORIGINAL SCIENTIFIC PAPER

Integration of Bivariate Logistic Regression Models and Decision Trees to Explore the Relationship between Socio-Demographic and Anthropometric Factors with the Incidence of Hypertension and Diabetes in Prospective Athletes

A'yunin Sofro[1], Danang Ariyanto[1], Junaidi Budi Prihanto[2], Dimas A. Maulana[1], Riska W. Romadhonia[1], Asri Maharani[3]

[1]Universitas Negeri Surabaya, Department of Mathematics, Surabaya, Indonesia, [2]Universitas Negeri Surabaya, Department of Sport Education, Surabaya, Indonesia, [3]University of Manchester, Department of Nursing, Manchester Area, United Kingdom

Abstract

Hypertension and diabetes are two medical conditions that are often associated with athletes' health. Hypertension or high blood pressure is a condition where the blood pressure in the arteries becomes too high. Meanwhile, diabetes is a condition where the body cannot produce or use insulin properly, thereby causing high blood sugar levels. Athletes' health is very important because they need optimal physical conditions to be able to compete effectively. Hypertension and diabetes can affect athletes' health and their performance. Socio-demographic and anthropometric factors are believed to play an important role in the development of both conditions. The aim of this study is to determine the relationship between socio-demographic and anthropometric factors on the incidence of hypertension and diabetes in prospective athletes in athletics and determine whether prospective athletes pass the initial screening process. This study integrates bivariate logistic regression models and decision trees to analyze data collected from 200 athlete selection participants. The univariate logistic regression model showed that waist circumference, father's occupation, and salary category 2 had a significant influence on hypertension, while BMI had a significant influence on diabetes. Meanwhile, the bivariate logistic regression model found that BMI and salary category 2 had a significant effect on hypertension. The optimal classification tree was formed using variables such as BMI, Salary Category 2, Hypertension, and Diabetes. The accuracy of the prediction data was 72%, indicating that the optimal tree is well-formed and suitable for classifying athletes' data. This study concludes that there is a significant relationship between sociodemographic and anthropometric factors and the incidence of hypertension and diabetes in prospective athletes. This study provides valuable insight into physiological adaptation, fitness, recovery, and other factors that influence athlete performance.

Keywords: *athletes, bivariate, decision tree, diabetes, hypertension, logistic regression*

Correspondence:

Danang Ariyanto
Universitas Negeri Surabaya, Department of Mathematics, Ketintang, 60231 Surabaya, Indonesia
E-mail: danangariyanto@unesa.ac.id

Introduction

Intense training in athletes at risk of hypertension and diabetes can have a significant impact on their health. The study by Fagard (2005) showed that hypertension has a significant effect on athlete performance, especially in athletic sports such as marathon running. On the one hand, intense exercise can help reduce the risk of developing both conditions by improving cardiovascular conditions, increasing insulin sensitivity, and aiding in maintaining a healthy weight (Luan et al., 2019). However, some aspects need to be carefully considered.

Training too intensely and too often without enough recovery time can increase the risk of injury and overtraining, which can negatively affect an athlete's health (Impellizzeri et al., 2020). In addition, intense training can also cause a temporary increase in blood pressure during physical activity, which is normal but needs to be carefully monitored, especially if the athlete is at risk of hypertension (Alpsoy, 2020).

Additionally, athletes who have insulin resistance or diabetes need to understand how intense training can affect their blood sugar levels. This may require more careful diet planning and proper dosing of insulin or medications (Muhtar & Lengkana, 2021).

In this context, careful management of intense exercise is essential. This includes regular health monitoring, adjustment of exercise intensity as per individual needs, appropriate diet planning, stress management, and consultation with a medical team experienced in treating athletes with these health risks (Jimenez et al., 2007). With a holistic and professional approach, intense training can still be a powerful tool in improving an athlete's health and performance while managing the risk of hypertension and diabetes (Trojian et al., 2022).

Hypertension, or high blood pressure, is a condition in which the blood pressure in the arteries persistently rises above normal levels (Fuchs & Whelton, 2020). In athletes, hypertension can be caused by overtraining, dehydration, and taking certain supplements. Hypertension in athletes can lead to serious health problems, including the risk of heart disease and stroke (Basilico, 1999). Therefore, it is important for athletes to keep their blood pressure within the normal range by paying attention to adequate fluid and electrolyte intake and avoiding overexertion (Sathish et al., 2012; Berge et al., 2015).

Diabetes is a condition in which the body cannot produce or use insulin properly, leading to elevated blood sugar levels (Mukhtar et al., 2020). In athletes, type 2 diabetes, which is often caused by being overweight and lack of physical activity, can be managed by adopting a healthy diet and regular physical exercise (Kurniawati, 2019). However, athletes also need to pay attention to their blood sugar levels during training and competition, as low or high blood sugar levels can affect their performance (Siwi et al., 2017). Diabetes can affect an athlete's health by causing disruptions to the nervous system and blood circulation, resulting in fatigue and decreased endurance (Mileva & Zaidell, 2022). A study published in the the Sports Medicine journal showed that diabetes can affect athletes' performance by reducing the body's ability to produce energy effectively during physical activity (Burke & Hawley, 2018; Absil et al., 2019; Scott et al., 2019).

Hypertension and diabetes are two very common medical conditions worldwide and both are considered significant global public health problems (Zhou et al., 2021). Socio-demographic and anthropometric factors are believed to play an important role in the development of both conditions (Manios et al., 2019). Overall, it is suspected that socio-demographic factors, such as age, gender, education, and socioeconomic status, and anthropometric factors are vital to a person's risk of developing hypertension and diabetes (Dey et al., 2022). Several studies have shown that people who are older, male, less educated, and have low socioeconomic status have a higher risk of developing hypertension and diabetes (Bovet et al., 2002; Ford, 2005).

Anthropometric factors such as body weight, body mass index (BMI), waist circumference, and height are also associated with the risk of hypertension and diabetes. A study conducted by Misra et al. (2007) has shown that people who are overweight or obese, have a higher BMI, larger waist circumference, and lower height have a higher risk of developing hypertension and diabetes.

In addition, several studies have examined the relationship between hypertension and diabetes with socio-demographic and anthropometric factors in the general population (Mackenbach et al., 2000; Lim et al., 2012; Schrier et al., 2012). According to the results of these studies, it appears that there is a significant relationship between socio-demographic and anthropometric factors with the risk of hypertension and diabetes in the general population, but there has never been any study related to this in the field of sports, especially in athletics. In addition, in terms of methods of approaching the problem, the integration of bivariate logistic regression models and decision trees is used, which have not been combined by many researchers. The novelty in this study lies in 2 points, namely the application of science in the field of sports, especially athletic sports, and the integration of bivariate logistic regression models and decision trees.

Methods

The data obtained is primary data from 200 athlete selection participants at the University of Surabaya and the East Java Indonesian National Sports Committee. In the selection process, data collection was carried out on prospective athletes through 2 processes. In the first screening, two observations were made, namely socio-demographics and anthropometry.

Procedure

Socio-demographic observations involve collecting data on the social and demographic characteristics of individuals or groups. This includes factors such as age (divided into 2 categories, <21 years and ≥21 years), gender (divided into 2 categories, female and male), father's education (divided into 2 categories, school and college), mother's education (divided into 2 categories, school and college), father's occupation (divided into 2 categories, formal and informal), mother's occupation (divided into 2 categories, formal and informal), and salary (divided into 3 categories, <3 million, 3-6 million, and >6 million). To conduct socio-demographic observations, survey methods, interviews, direct observation, and secondary data analysis were used. Anthropometric observations on athletes were conducted with the aim of understanding body proportions and physical characteristics that affect athletes' abilities in certain sports, includings factors such as height (divided into 2 categories, <170 cm and ≥170 cm), weight (divided into 2 categories, <60 kg and ≥60 kg), body mass index (divided into 2 categories, <25 and ≥25), and waist (divided into 2 categories, <85 cm and ≥85 cm).

Measurement

In the second screening process, 2 measurements regarding hypertension and diabetes were taken. Hypertension measurement was carried out by measuring the blood pressure of prospective athletes. If a prospective athlete has blood pressure above 120/80 mmHg, then they are indicated as having hypertension. Diabetes measurements include blood sugar tests and diabetes symptoms. After the observation and data collection process were carried out, the next stage was the data analysis process as a determinant of the initial screening results of prospective athletes to decide whether they pass or fail.

Statistics

In accordance with the problems to be studied, this study uses an integrated method in the problem-solving approach, namely combining bivariate logistic regression analysis and decision trees. First, bivariate logistic regression was used to observe the relationship between the incidence of hypertension and diabetes with socio-demographic and anthropometric factors. This method can be used to identify the effect of predictor variables on the dependent variable as well as to make predictions of the dependent variable category based on the values of the predictor variables (Liao, 1994; Sloane & Morgan, 1996; Hosmer et al., 2000).

A decision tree is a decision-making method in the form of a tree structure used to solve classification or regression problems (James et al., 2013). In a decision tree, each node on the tree represents a decision or prediction, while each branch of the tree represents a rule or condition that separates the data into smaller subgroups. Decision trees can be used to explore existing patterns or rules in data as well as to make predictions or classify new data. Both bivariate logistic regression and decision trees can be used to make predictions or classify data based on given predictor variables. However, the main difference between the two is the form of representation of the analysis results. Bivariate logistic regression produces a mathematical equation in the form of a logistic function used to estimate the probability of success or failure of the dependent variable based on the predictor variables. Meanwhile, decision tree produces a tree structure that describes the rules or conditions that must be met to make a prediction or classification (Kavakiotis et al., 2017; Rabbani et al., 2022; Ueshima et al., 2003). Integrating bivariate logistic regression analysis and decision trees is hoped to speed up the initial screening process for prospective athletic athletes with more accurate results.

Equations Bivariate Logistic Regression

Bivariate logistic regression is a statistical analysis technique used to analyze the relationship between two categorical or binary variables. In bivariate logistic regression analysis, the dependent variable is a binary or categorical variable, while the independent variable can be either categorical or continuous. The results of the analysis can be expressed in the form of odds ratio, which is a measure of the effect of the independent variable on the dependent variable.

Estimation of the parameter value is conducted using the Maximum Likelihood Estimation method by maximizing the natural logarithm of the likelihood function with the following formula (Hosmer et al., 2000):

$$\frac{\partial \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \left\{ \frac{y_{11}}{\pi_{11}} \frac{\partial \pi_{11}}{\partial \boldsymbol{\beta}} + \frac{y_{10}}{\pi_{10}} \frac{\partial \pi_{10}}{\partial \boldsymbol{\beta}} + \frac{y_{01}}{\pi_{01}} \frac{\partial \pi_{01}}{\partial \boldsymbol{\beta}} + \frac{y_{00}}{\pi_{00}} \frac{\partial \pi_{00}}{\partial \boldsymbol{\beta}} \right\}$$

n is the number of independent random samples, while are binomially distributed bivariate random variables with probability values of .

Decision Tree

Decision trees are one of the popular data analysis techniques in data science and computer science. This technique is used to generate prediction models based on existing data (Cheung Chiu & Webb, 1998). In decision tree analysis, data is divided into smaller subsets and then decisions are made based on a series of questions or rules applied to each subset. This method is used to predict the value of the dependent variable based on the independent variables.

The method used in decision trees is CART (Classification and Regression Trees). CART can be used to build decision trees for classification and regression. The CART algorithm goes through three stages, namely the formation of classification tree, pruning of the classification tree, and determination of the optimum classification tree.

Integrated Methods

The analysis procedure in this study is to integrate the bivariate logistic regression model and Decision Tree. The stages of analysis carried out were describing the data of respondents who participated in the athlete selection process by conducting descriptive statistical analysis. Then, modelling Bivariate Logistic Regression Analysis was carried out to determine Anthropometric and Socio-demographic Variables that have a significant influence on indications of hypertension and diabetes, where the level of significance is 0.05. Variable results obtained were then inputted to perform decision tree analysis with the CART method. To calculate the classification accuracy of the decision tree, the analysis was carried out using R studio software.

Results

An overview of the characteristics of respondents who participated in the selection of athletes at University of Surabaya and the Indonesian National Sports Committee of East Java will be provided based on anthropometric and sociodemographic factors that influence the indication of hypertension and diabetes. The data is presented in Table 1.

Table 1 shows a related percentage between anthropometric and sociodemographic factors with early screening for hypertension (high blood pressure) and diabetes (high sugar levels). First, the initial screening for hypertension in respondents with a height above 170 cm had a greater hypertension percentage of 17.50%, leading by 0.50% compared to those with a height less than 170 cm. On the other hand, during the initial screening for diabetes, respondents with a height below 170 cm showed a greater percentage of 20.50% compared to respondents above 170 cm with that of around 17%.

Table 1. Percentage of respondents' characteristics during initial screening

Variable		Hypertension (%)		Diabetes (%)	
		Yes	No	Yes	No
Height	< 170 cm	17.0	30.5	20.5	27.0
	≥170 cm	17.5	35.0	17.0	35.5
Weight	< 60 kg	18.0	35.5	21.0	32.5
	≥ 60 kg	16.5	30.0	16.5	30.0
Body Mass Index (BMI)	< 25	11.5	14.5	15.0	11.0
	≥ 25	23.0	51.0	22.5	51.5
Waist	< 85 cm	15.5	17.0	14.0	18.5
	≥ 85 cm	19.0	48.5	23.5	44.0
Age	< 21 years	22.5	40.0	24.5	38.0
	≥ 21 years	12.0	25.5	13.0	24.5
Gender	Female	15.5	36.0	18.5	33.0
	Male	19.0	29.5	19.0	29.5
Father's Education	School	14.5	24.5	16.5	22.5
	College	20.0	41.0	21.0	40.0
Mother's Education	School	14.0	25.0	17.0	22.0
	College	20.5	40.5	20.5	40.5
Father's Occupation	Formal	14.5	33.5	17.5	30.5
	Informal	20.0	32.0	20.0	32.0
Mother's Occupation	Formal	14.0	19.5	14.5	19.0
	Informal	20.5	46.0	23.0	43.5
Salary	< Rp 3.000.000	19.0	29.5	20.5	28.0
	3.000.000-6.000.000	13.5	28.5	15.0	27.0
	> Rp 6.000.000	2.0	7.5	2.0	7.5

Univariate Logistic Regression

Modeling logistic regression with one predictor variable is used to find out how much influence each predictor variable has on each response variable partially. The steps taken include making an estimate of the parameters for each predictor variable in the partial model.

Variable of Hypertension

The modeling results of independent variables on the hypertension response variable demonstrates that not every predictor variable exhibits a significant effect on the indication of hypertension suffered by athletes. The test results are presented in Table 2.

Table 2. Univariate logistic analysis for the hypertension response variable

Variable	Estimate	Std. Error	Z-value	P-value	Exp(Coef.)
Intercept	-1.232	0.441	-2.797	0.005**	0.292
Height	-0.110	0.366	-0.300	0.764	0.896
Weight	-0.068	0.352	-0.193	0.847	0.934
Body Mass Index (BMI)	0.289	0.398	0.725	0.468	1.335
Waist	0.737	0.363	2.030	0.042*	2.090
Age	0.461	0.352	1.308	0.191	1.585
Gender	-0.388	0.350	-1.107	0.269	0.679
Father's Education	0.180	0.334	0.539	0.590	1.197
Mother's Education	-0.073	0.365	-0.200	0.842	0.930
Father's Occupation	-0.703	0.333	-2.109	0.035*	0.495
Mother's Occupation	0.312	0.361	0.862	0.389	1.366
Salary_1	0.387	0.349	1.110	0.267	1.473
Salary_2	0.753	0.376	2.001	0.045*	2.124

Legend : *significant level 5%, **significant level 1%

In the univariate logistic regression analysis model with independent variable of hypertension, there are significant dependent variables, namely waist, father's occupation, salary category 2.

Variable of Diabetes

The univariate logistic regression modeling between independent variables and diabetes response variable is shown in Table 3.

Table 3. Logistic regression model with diabetes response variable

Variable	Estimate	Std. Err.	Z-value	P-value	Exp(Coef.)
Intercept	-1.238	0.435	-2.843	0.005**	0.290
Height	0.224	0.356	0.630	0.529	1.251
Weight	0.002	0.347	0.005	0.996	1.002
Body Mass Index (BMI)	1.095	0.390	2.810	0.005*	2.989
Waist	-0.028	0.364	-0.078	0.938	0.972
Age	0.383	0.340	1.128	0.259	1.467
Gender	-0.181	0.346	-0.523	0.601	0.834
Father's Education	0.370	0.326	1.134	0.257	1.447
Mother's Education	0.331	0.356	0.930	0.352	1.393
Father's Occupation	-0.267	0.320	-0.835	0.404	0.766
Mother's Occupation	0.105	0.358	0.294	0.769	1.111
Salary_1	-0.277	0.344	-0.806	0.420	0.7582
Salary_2	0.307	0.368	0.834	0.405	1.360

Legend : *significant level 5%, **significant level 1%

In the univariate logistic regression analysis model with independent variable of diabetes, body mass index (BMI) serves as a significant dependent variable.

Bivariate Logistic Regression

To conduct bivariate regression analysis, it is necessary to assume the existence of a relationship or correlation between the response variables.

According to the table above, it can be concluded that there is a significant relationship between the two response variables.

Based on the table above, it is worth noting that BMI and Salary coefficients have an influence on the hypertension response variable. This is proven by the test statistical values on the independent variable being -3.400 and 2.233 respectively, which are smaller than). As specified by the results of the bivariate logistic regression analysis, BMI and Salary category 2 are the variables with a significant influence. Therefore, these variables will be included in the independent variables in the decision tree analysis.

Decision Tree

Out of 4 predictor variables which comprises variables of BMI, Salary category 2, Hypertension, and Diabetes in the classification tree, variables provide optimum classification results. The formed optimum classification tree is illustrated in Figure 1.

Table 4. Correlation test results between the two response variables

Relationship between response variables		p-value	Summary
Y1 and Y2	18.835	0.000	Reject H0

Table 5. Bivariate Binary Logistic Regression Analysis of Each Predictor Variable

Coefficient	Estimate	Std Error	Z value	P-value
(Intercept):1	0.548	0.151	3.626	0.000***
(Intercept):2	0.620	0.153	4.052	0.000***
(Intercept):3	0.956	0.258	3.713	0.000***
BMI:1	-0.145	0.221	-0.659	0.510
BMI:2	-0.752	0.221	-3.400	0.001***
Waist:1	-0.395	0.210	-1.884	0.060
Waist:2	-0.011	0.214	-0.052	0.959
Father's Occupation:1	0.284	0.189	1.507	0.132
Father's Occupation:2	0.140	0.189	0.740	0.459
Salary_2:1	-0.446	0.200	-2.233	0.026*
Salary_2:2	-0.334	0.2018	-1.657	0.098

Legend : *significant level 5%, **significant level 1%, *** significant level 0.1%

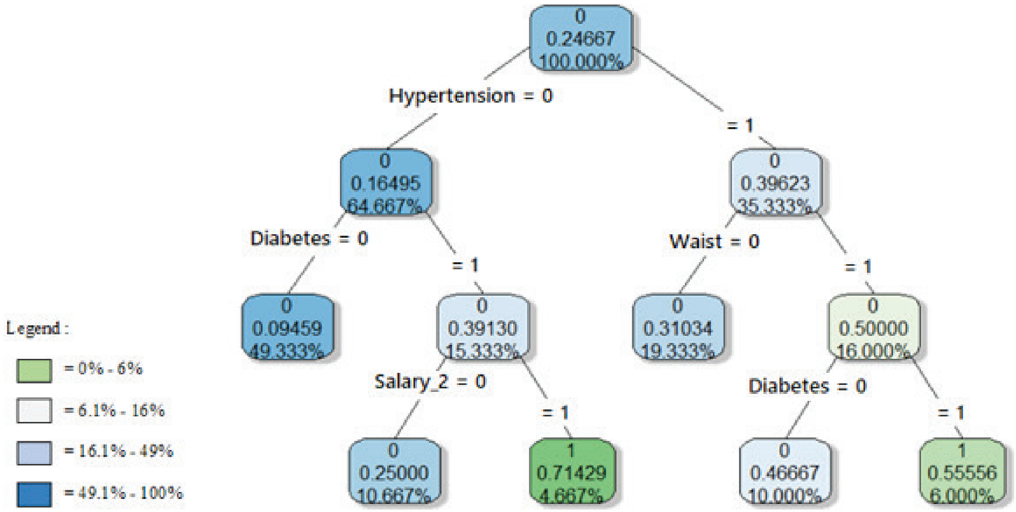

FIGURE 1. Optimum Classification Tree

The tree plot above exhibits the variables used to build the tree. The selected tree contains 4 variables with 6 splits. These variables are Hypertension, Diabetes, Waist Circumference and Salary category 2. Based on the decision tree structure in Figure node 3, after experiencing splitting, node 3 branches into two subsequent nodes. The first one is labeled as node 4 with no indication of diabetes, leading to it becoming the terminal node. Meanwhile, the second one is labeled as node 5 which represents indicated diabetes, further branching into node 8, namely salary category 2 (less than 6,000,000) and node 9, denoting salary category 2 (more than 6,000,000). Similarly, at node 3, it branches into two subsequent nodes. One, node 6 represents waist circumference of less than 80 cm, serving as a terminal node. The other is node 7 with a waist circumference of more than 80 cm which is divided into node 10, signifying those without diabetes, and node 11 with diabetes.

The next step is to calculate the classification accuracy of the CART tree obtained. The classification accuracy results for the learning data are presented in Table 6.

Table 6. Classification accuracy results

Athletes	Failed	Succeed
Fail	31	10
Succeed	4	5

The calculation of accuracy, sensitivity and specificity for the testing data is as follows:

$$\text{Accuracy} = \left(\frac{31+5}{50}\right)x100\% = 72\%$$

$$\text{Precision} = \left(\frac{31}{31+10}\right)x100\% = 75.61\%$$

$$\text{Sensitivity} = \left(\frac{31}{31+4}\right)x100\% = 88.57\%$$

$$\text{Specificity} = \left(\frac{5}{5+10}\right)x100\% = 33.33\%$$

Based on the conducted calculations, the accuracy value of the prediction data is 72 percent. Therefore, it is arguable that the optimal tree formed is good and suitable for classifying new data. Alongside this number, the precision, sensitivity and specificity values are also obtained, amounting to 75.61%, 88.57% and 33.33% respectively.

Discussion

A research conducted by Ruspriyanty and Sofro (2018) examined hypertension using Logistic Regression and Probit. Such research that integrates bivariate logistic regression models and decision trees in this study is a novel approach, especially in the context of athletics, and contributes to the methodological advances in this field. The findings of the bivariate logistic regression analysis and decision tree modeling provide valuable insights into the relationship between socio-demographic and anthropometric factors with the risk of hypertension and diabetes in prospective athletes. The significant relationship between both factors and the risk of hypertension and diabetes in athletes underscores the importance of taking them into account during the athletes selection and training to mitigate health risks and optimize their performance.

The results of the analysis revealed that hypertension was influenced by socio-demographic factors, namely the father's occupation and income. Apart from that, anthropometric factors also contribute a significant influence on hypertension, namely through waist circumference. This finding is in line with the research conducted by Park and Kim (2018) on 1,032 adults aged 20-80 years which found that the waist circumference variable had a significant effect on hypertension.

Another research conducted by Chen et al. (2018) on 211,833 teenagers over 20 years old in China concluded that there was a linear association between baseline BMI and the risk of developing diabetes which increased with each kg/m2 of BMI and was associated with 23% (95% CI 1.22% to 1.24%) higher risk of incident diabetes. The risk of incident diabetes grew by 35% (95% CI 1.29% to 1.40%) for each kg/m2 increase of BMI in the group of 20–30 years old and by 31% (95% CI 1.29% to 1.33%) in the group of 30–40 years old.

The odds ratio values obtained from the logistic regression models offer specific insights into the magnitude of influence exerted by different socio-demographic and anthropometric variables on the risk of hypertension and diabetes in athletes, providing valuable information for targeted interventions and support. The decision tree analysis, particularly using the CART method, offers a practical and effective approach to predicting the risk of hypertension and diabetes based on the identified socio-demographic and anthropometric factors which can be valuable for developing tailored intervention strategies. The integration of bivariate logistic regression model and decision tree was used to determine the relationship. The optimal classification tree formed from the data showed that the diabetes variable had the most significant impact on the participants' success as athletes. The accuracy of the prediction data was found to be 72%, indicating that the optimal tree formed was good and suitable for classifying athletes' data. This research provides valuable insights into physiological adaptation, fitness, recovery, and other factors that influence athletic performance. The research findings contribute to the existing body of knowledge on the relationship between socio-demographic and anthropometric factors and the risk of hypertension and diabetes, particularly in the context of athletes, and provide a foundation for further research and practical applications in sports medicine and athlete management. The implications of the research findings extend to the development of evidence-based guidelines for athlete selection, training, and support programs, focusing on mitigating the risk of hypertension and diabetes through targeted interventions based on socio-demographic and anthropometric considerations.

Acknowledgements
The authors express their gratitude to the Directorate General of Higher Education, Ministry of Education and Culture for the basic research grant that provided funding to support this research.

Conflict of Interest
The author declares that there is no conflict of interest.

Received: 25 November 2023 | **Accepted:** 13 January 2024 | **Published:** 01 February 2024

References

Absil, H., Baudet, L., Robert, A., & Lysy, P. A. (2019). Benefits of physical activity in children and adolescents with type 1 diabetes: a systematic review. *Diabetes Research and Clinical Practice*, *156*, 107810.

Alpsoy, Ş. (2020). Exercise and hypertension. *Physical Exercise for Human Health*, 153–167.

Basilico, F. C. (1999). Cardiovascular disease in athletes. *The American Journal of Sports Medicine*, *27*(1), 108–121.

Berge, H. M., Isern, C. B., & Berge, E. (2015). Blood pressure and hypertension in athletes: a systematic review. *British Journal of Sports Medicine*, *49*(11), 716–723. https://doi.org/10.1136/bjsports-2014-093976

Bovet, P., Ross, A. G., Gervasoni, J.-P., Mkamba, M., Mtasiwa, D. M., Lengeler, C., Whiting, D., & Paccaud, F. (2002). Distribution of blood pressure, body mass index and smoking habits in the urban population of Dar es Salaam, Tanzania, and associations with socioeconomic status. *International Journal of Epidemiology*, *31*(1), 240–247.

Burke, L. M., & Hawley, J. A. (2018). Swifter, higher, stronger: what's on the menu? *Science*, *362*(6416), 781–787.

Chen, Y., Zhang, X.-P., Yuan, J., Cai, B., Wang, X.-L., Wu, X.-L., … & Zhu, X.-H. (2018). Association of body mass index and age with incident diabetes in Chinese adults: a population-based cohort study. *BMJ Open*, *8*(9).

Cheung Chiu, B., & Webb, G. I. (1998). Using decision trees for agent modeling: improving prediction performance. *User Modeling and User-Adapted Interaction*, *8*, 131–152.

Dey, S., Mukherjee, A., Pati, M. K., Kar, A., Ramanaik, S., Pujar, A., … & Swaroop, N. (2022). Socio-demographic, Behavioural and Clinical determinants influencing Control of Diabetes and Hypertension in Urban Mysore, South India: A Mixed-Method Study. *Archives of Public Health, 80*(1), 234

Fagard, R. H. (2005). Effects of exercise, diet and their combination on blood pressure. *Journal of Human Hypertension*, *19*(3), S20–S24.

Ford, E. S. (2005). Risks for all-cause mortality, cardiovascular disease, and diabetes associated with the metabolic syndrome: a summary of the evidence. *Diabetes Care*, *28*(7), 1769–1778.

Fuchs, F. D., & Whelton, P. K. (2020). High blood pressure and cardiovascular disease. *Hypertension*, *75*(2), 285–292.

Hosmer, D. W., Lemeshow, S., & Cook, E. (2000). *Applied logistic regression 2nd edition.* New York: Jhon Wiley and Sons Inc.

Impellizzeri, F. M., Menaspà, P., Coutts, A. J., Kalkhoven, J., & Menaspà, M. J. (2020). Training load and its role in injury prevention, part I: back to the future. *Journal of Athletic Training*, *55*(9), 885–892.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). Springer.

Jimenez, C. C., Corcoran, M. H., Crawley, J. T., Hornsby Jr, W. G., Peer, K. S., Philbin, R. D., & Riddell, M. C. (2007). National athletic trainers' association position statement: management of the athlete with type 1 diabetes mellitus. *Journal of Athletic Training*, *42*(4), 536.

Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine learning and data mining methods in diabetes research. *Computational and Structural Biotechnology Journal*, *15*, 104–116.

Kurniawati, N. (2019). *Penerapan Konseling Aktivitas Fisik Dan Pola Makan Terhadap Perubahan IMT dan Kadar Glukosa Darah Pada Keluargadengan Diabetes Mellitus Tipe 2*. Skripsi, Universitas Muhammadiyah Magelang.

Liao, T. F. (1994). *Interpreting probability models: Logit, probit, and other generalized linear models* (Issue 101). Sage.

Lim, S. S., Vos, T., Flaxman, A. D., Danaei, G., Shibuya, K., Adair-Rohani, H., … & Andrews, K. G. (2012). A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet*, *380*(9859), 2224–2260.

Luan, X., Tian, X., Zhang, H., Huang, R., Li, N., Chen, P., & Wang, R. (2019). Exercise as a prescription for patients with various diseases. *Journal of Sport and Health Science*, *8*(5), 422–441.

Mackenbach, J. P., Cavelaars, A., Kunst, A. E., & Groenhof, F. (2000). Socioeconomic inequalities in cardiovascular disease mortality. An international study. *European Heart Journal*, *21*(14), 1141–1151.

Manios, Y., Karatzi, K., Moschonis, G., Ioannou, G., Androutsos, O., Lionis, C., & Chrousos, G. (2019). Lifestyle, anthropometric, socio-demographic and perinatal correlates of early adolescence hypertension: The Healthy Growth Study. *Nutrition, Metabolism and Cardiovascular Diseases*, *29*(2), 159–169.

Mileva, K. N., & Zaidell, L. (2022). Sport and Exercise Science and Health. In *Health studies: An introduction* (pp. 85–124). Springer.

Misra, A., Khurana, L., Vikram, N. K., Goel, A., & Wasir, J. S. (2007). Metabolic syndrome in children: current issues and South Asian perspective. *Nutrition*, *23*(11–12), 895–910.

Muhtar, T., & Lengkana, A. S. (2021). *Kesehatan Dalam Pendidikan Jasmani*. CV Salam Insan Mulia.

Mukhtar, Y., Galalain, A., & Yunusa, Ujej. (2020). A modern overview on diabetes mellitus: a chronic endocrine disorder. *European Journal of Biology*, *5*(2), 1–14.

Park, S. H., & Kim, S.-G. (2018). Comparison of hypertension prediction analysis using waist measurement and body mass index by age group. *Osong Public Health and Research Perspectives*, *9*(2), 45.

Rabbani, N., Kim, G. Y. E., Suarez, C. J., & Chen, J. H. (2022). Applications of machine learning in routine laboratory medicine: Current state and future directions. *Clinical Biochemistry*, *103*, 1–7.

Ruspriyanty, D. I., & Sofro, A. (2018). Analysis of Hypertension Disease using Logistic and Probit Regression. *Journal of Physics: Conference Series*, *1108*(1). https://doi.org/10.1088/1742-6596/1108/1/012054

Sathish, T., Kannan, S., Sarma, P. S., Razum, O., & Thankappan, K. R. (2012). Incidence of hypertension and its risk factors in rural Kerala, India: a community-based cohort study. *Public Health*, *126*(1), 25–32.

Schrier, A. C., Ujcic-Voortman, J. K., De Wit, M. A. S., Verhoeff, A. P., Kupka, R., Dekker, J., & Beekman, A. T. F. (2012). Ethnic differences in the association between cardiovascular risk factors and psychological distress in a population study in the Netherlands. *BMC Public Health*, *12*(1), 1–7.

Scott, S. N., Anderson, L., Morton, J. P., Wagenmakers, A. J. M., & Riddell, M. C. (2019). Carbohydrate restriction in type 1 diabetes: a realistic therapy for improved glycaemic control and athletic performance? *Nutrients*, *11*(5), 1022.

Siwi, T. P., Dieny, F. F., & Fitranti, D. Y. (2017). *Pengaruh Pemberian Diet Dengan Pengaturan Indeks Glikemik Dan Beban Glikemik Terhadap Kadar Glukosa Darah Atlet Sepak Bola Remaja*. Diponegoro University.

Sloane, D., & Morgan, S. P. (1996). An introduction to categorical data analysis. *Annual Review of Sociology*, *22*(1), 351–375.

Trojian, T., Colberg, S., Harris, G., Oh, R., Dixit, S., Gibson, M., Corcoran, M., Ramey, L., & Berg, P. V. (2022). American medical society for sports medicine position statement on the care of the athlete and athletic person with diabetes. *Clinical Journal of Sport Medicine*, *32*(1), 8–20.

Ueshima, H., Okayama, A., Saitoh, S., Nakagawa, H., Rodriguez, B., Sakata, K., Okuda, N., Choudhury, S. R., & Curb, J. D. (2003). Differences in cardiovascular disease risk factors between Japanese in Japan and Japanese-Americans in Hawaii: the INTERLIPID study. *Journal of Human Hypertension*, *17*(9), 631–639.

Zhou, B., Perel, P., Mensah, G. A., & Ezzati, M. (2021). Global epidemiology, health burden and effective interventions for elevated blood pressure and hypertension. *Nature Reviews Cardiology*, *18*(11), 785–802.