International Journal of Electrical and Computer Engineering (IJECE)
Vol. 14, No. 1, February 2024, pp. 854~860
ISSN: 2088-8708, DOI: 10.11591/ijece.v14i1.pp854-860

Predictive model for acute myocardial infarction in working-age population: a machine learning approach

Astrid Lorena Urbano-Cano[1], Diana Jimena López-Mesa[2], Rosa Elvira Alvarez-Rosero[3], Yeison Alberto Garcés-Gómez[4]
[1]Department of Biology, Faculty of Natural Sciences and Education, Universidad del Cauca, Popayán, Colombia
[2]Faculty of Engineering, Corporación Universitaria Comfacauca, Popayán, Colombia
[3]Department of Physiological Sciences, Faculty of Health Sciences, University of Cauca, Popayán, Colombia
[4]Faculty of Engineering and Architecture, Universidad Católica de Manizales, Manizales, Colombia

Article Info

Article history:

Received Jul 9, 2023
Revised Jul 4, 2023
Accepted Jul 17, 2023

Keywords:

Cardiovascular risk
Machine learning
Myocardial infarction
Prediction model
Random forest

ABSTRACT

Cardiovascular diseases are the leading cause of mortality in Latin America, particularly acute myocardial infarction (AMI), which is the primary cause of atherosclerotic cardiovascular morbidity. This study aims to develop a predictive model for the probability of AMI occurrence in the working-age population, based on atherogenic indices, paraclinical variables, and anthropometric measures. The research conducted a cross-sectional study involving 427 workers aged 40 years or older in Popayán, Colombia. Out of this population, 202 individuals were screened with a 95% confidence interval and a 5% error margin. Epidemiological, anthropometric, and paraclinical data were collected. A binary logistic regression model was employed to identify variables directly associated with the probability of AMI. Predictive classification models were generated using statistical software JASP and the programming language Python. During the training stage, JASP produced a model with an accuracy of 87.5%, while Python generated a model with an accuracy of 90.2%. In the validation stage, JASP achieved an accuracy of 93%, and Python reached 95%. These results establish an effective model for predicting the probability of AMI in the working population.

Corresponding Author:

Yeison Alberto Garcés-Gómez
Department of Electrical Engineering, Faculty of Engineering and Architecture, Universidad Católica de Manizales
Cra 23 No 60-63, Manizales, Colombia
Email: ygarces@ucm.edu.co

1. INTRODUCTION

Cardiovascular disease (CVD) is the leading cause of morbidity and mortality worldwide, affecting millions of people every year. CVD is largely preceded by atherosclerosis, which is a major precursor to the three most common vascular diseases: acute myocardial infarction (AMI), stroke, and peripheral arterial disease (PAD). Together, these diseases represent approximately one-third of all deaths worldwide [1], [2].

The AMI is defined as a necrosis in a clinical setting compatible with acute myocardial ischemia, which is commonly secondary to thrombotic occlusion of the coronary artery. Pain is a dominant characteristic symptom of this condition. It should be noted that the AMI is also based on the presence of myocardial damage detected by the elevation of cardiac biomarkers in the context of evidence of acute myocardial ischemia [3]. The AMI contributes more than 80% of cases of ischemic heart disease due to atherosclerosis, being more frequent in men than in women. Additional risk factors that can cause AMI include obesity, a high-calorie diet, smoking, type 2 diabetes mellitus, hypertension, and dyslipidemia, among others [4]. Atherosclerotic risk can be predicted by measuring atherogenic indices based on the lipid profile, which consist of the mathematical ratio or proportion between the levels of total cholesterol, triglycerides, high-density lipoprotein (HDL), or low-density lipoprotein (LDL) [5]. However, it is important to note that these indices are not explored in the working population of the region, which is necessary for the economic development of the territory. Therefore, it is necessary to include the work environment in this study to obtain more accurate results on the risk of developing AMI.

It is important to highlight that people with this pathology or a history of it have significantly reduced quality and life expectancy due to premature deaths and years of life lost due to disability. This represents a high health cost. Specifically, in individuals of working age, significant economic burdens are evidenced due to work disability, either due to absenteeism or loss of productivity. Therefore, it is necessary to create mechanisms that allow predicting the probability of occurrence of AMI in working individuals to create strategies that mitigate its effects within the work environment [6].

The use of autonomous learning algorithms or machine learning algorithms applied to the field of medicine is a novel topic. Currently, several studies in the literature report the use of algorithms such as artificial neural networks, case-based reasoning, Bayesian networks, decision trees, or k-means to support the diagnosis of diseases such as breast cancer, prostate cancer, cardiovascular diseases, hypertension, Parkinson's, infarctions, rheumatoid arthritis, among others, and the prediction of mortality or survival after cardiovascular events [7]. The use of these techniques becomes a fundamental support for healthcare personnel since some pathologies could be prevented and adequately treated before they present major complications, improving the survival chances of patients. Its application is also carried out in the emergency area of hospitals when diagnosing patient triage, reducing assessment times, and assigning the correct care shift [8]. Several studies related to this topic have been summarized in this document.

For pancreatic cancer, the study by [9] used neural networks to predict individual long-term survival of patients undergoing radical surgery for this type of cancer, with a performance of 79%. On the other hand, the research by [10] exposed the application of neural networks to recognize complex patterns for the prediction of advanced bladder cancer in patients undergoing radical cystectomy, breast cancer, and also prediction of survival after hepatic resection for colorectal cancer. The implemented model resulted in a disease prediction rate of 90.5% and individual survival prediction of 72% [11]–[13].

The research by [14] shows how classification algorithms such as logistic model tree (LMT), Bayesian networks, naive Bayes, J48, and naive Bayes simple were used for the diagnosis of pathologies in the Spine, to decide which is the best algorithm for the diagnosis of this disease. The results obtained during classification show that the LMT decision algorithm obtained a success rate of 85.48%. The Bayesian networks algorithm had a success rate of 80%; the naive Bayes algorithm correctly classified 248 instances with an absolute error of 2%, and finally, the naive Bayes simple algorithm correctly classified 241 instances of the 310, reaching the conclusion that the best decision algorithm is the LMT.

In [15], important results were obtained for the diagnosis of rheumatoid arthritis, as well as its categorization and potential application in personalized medicine for individuals affected by this disease. Computational models were designed for classification, among which are artificial neural networks that using 5 variables obtained a sensitivity of 92.3% with a specificity of 86.66%, and with Bayesian networks, a sensitivity of 92.3% and a specificity of 93.33% were achieved. Using artificial neural networks in [16], a model capable of recognizing 3 types of values: cirrhosis, non-cirrhosis, and non-identifiable with a success rate of almost 90% was obtained.

For the prognosis of bladder cancer mortality, [17] used seven learning methods to predict mortality at five years after radical cystectomy, including neural networks, radial basis function networks, extreme learning machine (ELM), regularized ELM (RELM), support vector machine (SVM), and the nearest neighbor classifier (K-NN). The results indicate that RELM achieves the highest prediction accuracy with 80%. In patients with stroke (ischemia and hemorrhage), for the prognosis of mortality 10 days after the event, the research by [18] applied neural networks to obtain a predictive model and achieved a sensitivity and accuracy of 87.8% for the hemorrhagic group and specificity of 75.9%, sensitivity of 85.9%, and accuracy of 80.9% for the ischemic group.

On the other hand, [19] show the training and testing of different neural networks for the diagnosis of myocardial infarction. The training and testing of several neural networks with different architectures were carried out for the diagnosis of infarction, based on the data from the Braunwald angina probability rating scale [20]. 40 networks were generated and tested in 5 experiments, of which the diagnostic accuracy was higher with the model of 5 electrocardiographic inputs plus troponin. Several of the networks designed for this case had a sensitivity and specificity close to 99%. In turn, [21] in their research expose different machine learning algorithms for the classification of breast cysts through thermographic images using artificial neural networks. Their results indicate a sensitivity of 78% and specificity of 88%. The overall efficiency of the system was 83%

In other areas of the medical field, artificial neural networks have been proposed to predict prolonged hospital stay for elderly patients in the emergency department, as well as prolonged stay in the intensive care unit and mortality, achieving a sensitivity of 62.5% and specificity of 96.6% for hospital stay and 82% for mortality prediction [22]–[24]. Previous research has shown that the use of autonomous learning algorithms is an effective tool for the diagnosis and prediction of disease onset, allowing the medical team to provide timely treatment and thus improving the patient's quality of life. Within the literature of the last five years, consulted by the authors of this study, there is no evidence of predictive models focused specifically on the probability of suffering from AMI related to atherogenic indices, anthropometric measures or paraclinical variables for the working population.

2. METHOD

The study is an observational, cross-sectional descriptive study, whose database started to be analyzed from January 2022. A cross-sectional study was conducted in 427 workers aged ≥40 years in the city of Popayán, from which 202 individuals were screened, considering a confidence interval of 95% and an error of 5%. Subsequently, epidemiological, clinical, and paraclinical data were collected, the latter from a peripheral blood sample after obtaining informed consent. All the questionnaires, procedures, and protocols were reviewed and approved by the Ethics Committee for Scientific Research at the University of Cauca; the guidelines used in their view were based on the bioethical principles established in the Helsinki in 1975 declaration and the parameters outlined in Resolution 8430 of the Colombian Ministry of Health in 1993.

Atherogenic indices (AI) were calculated: total-cholesterol-high-density-lipoprotein ratio (TC/HDL), low-density-lipoprotein-high-density-lipoprotein ratio (LDL/HDL), TC-HDL/HDL, TC-HDL, logarithmic triglycerides-to-high-density-lipoprotein (LOG(TG/HDL)), and TG/HDL. Cardiovascular risk was measured using the Framingham scale, using variables such as age, sex, systolic blood pressure, diabetes, smoker, total cholesterol levels, and HDL. Later, a correlation will be made between the atherogenic indices and the percentage of cardiovascular risk, estimating the risk of developing atherosclerosis [5] and the coronary risk according to the Framingham adjusted function, recommended by the Colombian guide [24]. The Framingham-adjusted scale (Framingham cardiovascular risk × 0.75) proposed recently by the clinical practice guide for the prevention, early detection, diagnosis, treatment, and follow-up of Dyslipidemia in Colombia [25], [26] will be used.

According to the purposes of this study, the occurrence of AMI was taken as the dependent variable, which is categorical and binary, with output values of "yes" or "no." The remaining variables that may be related to it are quantitative (cholesterol, triglycerides, glucose, HDL, LDL, very-low-density lipoprotein (VLDL), blood pressure, peripheral arterial disease (PAD), body mass index (BMI), abdominal perimeter (AP), age, ICI_CT/cHDL, cholesterol-high-density-lipoprotein CHOL-HDL/HDL) and some are categorical (sex, smoking, physical activity, marital status, education, race, origin, occupation, type of contract).

The analysis begins by determining the relationship between the different variables, so a logistic regression is performed given the nature of the dependent variable. Different combinations of variables are performed using the JASP statistical software until the most favorable result is obtained, where a relationship was found between the occurrence of acute myocardial infarction and the variables summarized in Table 1.

Table 1. The performance of risk variables related to AIM

Variable type	Name	Range or value
Quantitative	Cholesterol	(117 – 300) mg/dL
	Triglycerides	(88 – 424,7) mg/dL
	Blood glucose	(68 – 310) mg/dL
	BMI	(18 – 45,7) Kg/cm2
	AP	(42 – 162) cm
	PAD	(0,72 – 1,25)
	ICI_CT/cHDL	(2 – 16,22)
	CHOL-HDL/HDL	(1 – 15,22)
Categorical	Sex	M-F

3. RESULTS AND DISCUSSION

3.1. Logistic regression

According to the Akaike information criteria (AIC) and Bayesian information criteria (BIC) metrics, the alternative hypothesis (H1) has the lowest values, suggesting a significant relationship between the output and predictor variables as shown in Table 2. The McFadden R2 value is 0.339, indicating that it is a good model as its value falls within the range of 0.2-0.4. The *p*-value is 0.001, which, being less than 0.05, shows the good performance of the model. The Nagelkerke, Tjur; and Cox and Snell R2 values are very useful when comparing different models for the same data, where the model with the highest R2 values is considered the most appropriate. However, this is not the purpose of this analysis [27].

Table 2. The performance of logistic regression

Model	Deviance	AIC	BIC	df	X2	P	McFadden R2	Nagelkerke R2	Tjur R2	Cox &Snell R2
H0	278,425	280,425	283,734	201						
H1	184,031	204,031	237,113	192	94,395	<.001	0.339	0.499	0.477	0.373

According to the confusion matrix as shown in Table 3, out of the 202 entered data, 104 data that should have been classified as NO were correctly classified, and 77 that should have been classified as YES were correctly classified, showing an accuracy of 92.8%. The sensitivity is 83.7% and the specificity is 94.5% as seen in Table 4. These results show a good performance of the algorithm that also eliminates subjectivity and the need for highly trained medical personnel.

Table 3. The performance of confusion matrix obtained with logistics regression

	Predicted		
Observed	no	si	% Correct
no	104	6	94,545
si	15	77	83,696
Overall % Correct			89,604

Table 4. The performance of metrics

	Value
Sensitivity	0.837
Specificity	0.945
Precision	0.928

3.2. Machine learning algorithms

Given the results obtained in the logistic regression, it was possible to clearly identify the variables that will be part of the predictive model that is intended to be generated. For the construction of the model, the machine learning classification module is used, which proposes several methods, including boosting, decision tree, k-nearest neighbors, random forest, support vector machine [27]. Comparing the different metrics of the prediction methods related to machine learning classification, similarity between them is evident. The average accuracy for boosting, decision tree, and random forest is 87.5%, and their precision is 88.1%, while for k-nearest neighbors and support vector machine, the average accuracy is 85%, and their precision is 85.1%.

With these differences, although not significant, the selection is reduced to the methods: boosting, random forest, and decision tree, where the first two are based on decision tree ensembles, which may have an advantage over the decision tree technique, since according to the literature, this technique is not as robust, as a small change in the data can cause large changes in the final estimated tree [28]. The boosting and random forest methods differ in their training approach, since for random forest, each tree is trained individually with a slightly different random sample of the training data generated by bootstrapping, while in the boosting method, the trees are trained sequentially, so that each new tree tries to improve on the errors of the previous trees [29]. Given that the results obtained are similar among the tree-based ensemble techniques and due to its ease of interpretation, it was chosen to generate the predictive model using the random forest technique.

3.3. Random forest algorithm

The database was inputted into JASP software and by using the random forest method of the machine learning classification module, it was observed that with 14 trees each with 3 predictors per split and taking 129 training data (64%), 33 validation data (16%), and 40 evaluation data (20%), an accuracy of 87.5% is obtained for both YES and NO values. As for precision, it is 92.9% for YES values and 84.6% for NO values as shown in Tables 5 and 6, which represents very satisfactory results.

Table 5. Summary of the classification model generated

Trees	Predictors per split	n(Train)	n(Validation)	n(Test)	Validation accuracy	Test accuracy	OOB accuracy
14	3	129	33	40	0.909	0.875	0.893

Table 6. Confusion matrix of generated classification model

		Predicted	
		no	yes
Observed	no	22	1
	yes	4	13

Figure 1 shows the increase in accuracy for training data and validation data with respect to the number of trees. From 14 trees onwards, the accuracy for validation is 90.9% and slightly lower for the training data (87.5%). Finally, the ROC curves shown in Figure 2 indicate that both for the classification of YES and NO, their shapes approach the desired one, with an area under the curve of AUC of 0.863, that is, there is an 86.3% probability that the classification is correct.

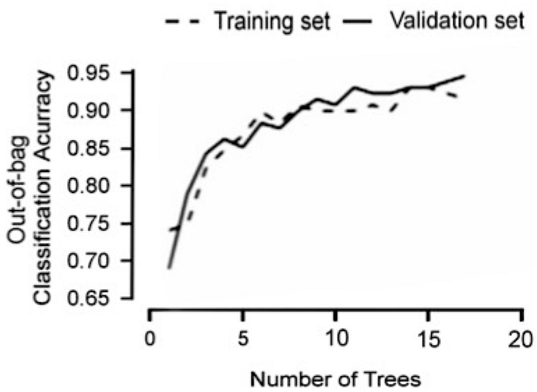

Figure 1. Relationship of accuracy with respect to the number of trees

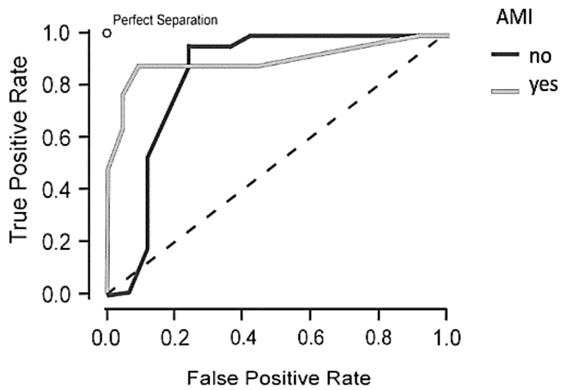

Figure 2. ROC curves of the classification model of the generated model

4. CONCLUSION

For the first time in the municipality of Popayan, a model has been established that allows predicting the probability of suffering an acute myocardial infarction in a population over the fourth decade of life with occupational activity. The mentioned model achieves a prediction accuracy of 95%. This result will help reduce the incidence and mortality rates of the study population, improving their quality of life.

The predictive model was generated in two computational tools: the statistical software JASP and the programming language Python, where the latter had better performance with a 2% difference in accuracy. The obtained classification model showed satisfactory results in its accuracy, where in the training phase, this value was 90.2% while in the validation phase, it increased to 95%. In the future, a cardiovascular risk calculator could be generated to be used in a simple and non-invasive way in clinical practice.

ACKNOWLEDGEMENTS

Colfuturo for the financing of the project: Social determination of health in formal workers in the city of Popayán: a territorial reading. Financed through call 002 Doctorate for National Teachers. Specialization in Applied Statistics from the Catholic University of Manizales. Biology and Physiological Science Departments of the University of Cauca. We are also indebted to all the volunteers who participated in the study. Finally, we acknowledge the collaboration of the applied human genetics group of the University of Cauca.

REFERENCES

[1] A. Nitsa, M. Toutouza, N. Machairas, A. Mariolis, A. Philippou, and M. Koutsilieris, “Vitamin D in cardiovascular disease,” *In Vivo*, vol. 32, no. 5, pp. 977–981, 2018, doi: 10.21873/invivo.11338.

[2] J. Soppert, M. Lehrke, N. Marx, J. Jankowski, and H. Noels, "Lipoproteins and lipids in cardiovascular disease: from mechanistic insights to therapeutic targeting," *Advanced Drug Delivery Reviews*, vol. 159, pp. 4–33, 2020, doi: 10.1016/j.addr.2020.07.019.
[3] K. Thygesen *et al.*, "Fourth universal definition of myocardial infarction (2018)," *Circulation*, vol. 138, no. 20, Nov. 2018, doi: 10.1161/cir.0000000000000617.
[4] B. Ibanez *et al.*, "2017 ESC guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation," *European Heart Journal*, vol. 39, no. 2, pp. 119–177, Jan. 2018, doi: 10.1093/eurheartj/ehx393.
[5] K. De la Torre-Cisneros, Z. Acosta-Rodríguez, and V. Aragundi-Intriago, "Clinical utility of atherogenic indices for cardiovascular risk assessment: a clinical laboratory approach (in Spain)," *Dominio de las Ciencias*, vol. 5, no. 3, Jul. 2019, doi: 10.23857/dc.v5i3.924.
[6] H. J. Warraich, L. A. Kaltenbach, G. C. Fonarow, E. D. Peterson, and T. Y. Wang, "Adverse change in employment status after acute myocardial infarction," *Circulation: Cardiovascular Quality and Outcomes*, vol. 11, no. 6, Jun. 2018, doi: 10.1161/CIRCOUTCOMES.117.004528.
[7] J. C. de J. Montero Rodríguez, R. Roshan Biswal, and E. S. de la Cruz, "State-of-the-art machine learning algorithms for disease diagnosis (in Spain)," *Research in Computing Science*, vol. 148, no. 7, pp. 455–468, Dec. 2019, doi: 10.13053/rcs-148-7-34.
[8] T. Hastie, R. Tibshirani, G. James, and D. Witten, "An introduction to statistical learning (2nd ed.)," *Springer texts*, vol. 102, 2021.
[9] D. Ansari, J. Nilsson, R. Andersson, S. Regnér, B. Tingstedt, and B. Andersson, "Artificial neural networks predict survival from pancreatic cancer after radical surgery," *The American Journal of Surgery*, vol. 205, no. 1, pp. 1–7, Jan. 2013, doi: 10.1016/j.amjsurg.2012.05.032.
[10] A. M. Vukicevic, G. R. Jovicic, M. M. Stojadinovic, R. I. Prelevic, and N. D. Filipovic, "Evolutionary assembled neural networks for making medical decisions with minimal regret: application for predicting advanced bladder cancer outcome," *Expert Systems with Applications*, vol. 41, no. 18, pp. 8092–8100, Dec. 2014, doi: 10.1016/j.eswa.2014.07.006.
[11] I. Saritas, "Prediction of breast cancer using artificial neural networks," *Journal of Medical Systems*, vol. 36, no. 5, pp. 2901–2907, Oct. 2012, doi: 10.1007/s10916-011-9768-0.
[12] L. Spelt, J. Nilsson, R. Andersson, and B. Andersson, "Artificial neural networks – a method for prediction of survival following liver resection for colorectal cancer metastases," *European Journal of Surgical Oncology (EJSO)*, vol. 39, no. 6, pp. 648–654, Jun. 2013, doi: 10.1016/j.ejso.2013.02.024.
[13] E. S. Wise, K. M. Hocking, and C. M. Brophy, "Prediction of in-hospital mortality after ruptured abdominal aortic aneurysm repair using an artificial neural network," *Journal of Vascular Surgery*, vol. 62, no. 1, pp. 8–15, Jul. 2015, doi: 10.1016/j.jvs.2015.02.038.
[14] N. V. R. Pérez, M. L. Estrada, and A. M. D. A. Tovar, "Application of artificial intelligence methods in the medical area (in Spain)," *Pistas Educativas*, vol. 35, no. 111, pp. 124–130, 2015.
[15] F. A. González, "Computational learning models in rheumatology (in Spain)," *Revista Colombiana de Reumatología*, vol. 22, no. 2, pp. 77–78, Jun. 2015, doi: 10.1016/j.rcreu.2015.06.001.
[16] V. M. Bostan and B. Pantelimon, "Creating a model based on artificial neural network for liver cirrhosis diagnose," in *2015 9th International Symposium on Advanced Topics in Electrical Engineering (ATEE)*, May 2015, pp. 295–298, doi: 10.1109/ATEE.2015.7133783.
[17] G. Wang, K.-M. Lam, Z. Deng, and K.-S. Choi, "Prediction of mortality after radical cystectomy for bladder cancer by machine learning techniques," *Computers in Biology and Medicine*, vol. 63, pp. 124–132, Aug. 2015, doi: 10.1016/j.compbiomed.2015.05.015.
[18] M. Alsalamah, S. Amin, and J. Halloran, "Diagnosis of heart disease by using a radial basis function network classification technique on patients' medical records," in *2014 IEEE MTT-S International Microwave Workshop Series on RF and Wireless Technologies for Biomedical and Healthcare Applications (IMWS-Bio2014)*, Dec. 2014, pp. 1–4, doi: 10.1109/IMWS-BIO.2014.7032401.
[19] J. J. S. Díaz, J. J. D. Fernández, and E. G. Guerrero, "Automatic diagnosis of acute coronary syndrome using a multiagent system based on neural networks (in Spain)," *Revista Colombiana de Cardiología*, vol. 24, no. 3, pp. 255–260, May 2017, doi: 10.1016/j.rccar.2016.11.010.
[20] R. Villar *et al.*, "Scales in internal medicine: cardiology (in Spain)," *Galicia Clin*, vol. 31, no. 1, pp. 31–36, 2019.
[21] M. A. de Santana *et al.*, "Breast cancer diagnosis based on mammary thermography and extreme learning machines," *Research on Biomedical Engineering*, vol. 34, no. 1, pp. 45–53, Mar. 2018, doi: 10.1590/2446-4740.05217.
[22] C. P. Launay, H. Rivière, A. Kabeshova, and O. Beauchet, "Predicting prolonged length of hospital stay in older emergency department users: use of a novel analysis method, the artificial neural network," *European Journal of Internal Medicine*, vol. 26, no. 7, pp. 478–482, Sep. 2015, doi: 10.1016/j.ejim.2015.06.002.
[23] R. Houthooft *et al.*, "Predictive modelling of survival and length of stay in critically ill patients using sequential organ failure scores," *Artificial Intelligence in Medicine*, vol. 63, no. 3, pp. 191–207, Mar. 2015, doi: 10.1016/j.artmed.2014.12.009.
[24] O. M. Muñoz V, Á. J. Ruiz Morales, A. Mariño Correa, and M. M. Bustos C., "Concordancia entre los modelos de SCORE y Framingham y las ecuaciones AHA/ACC como evaluadores de riesgo cardiovascular," *Revista Colombiana de Cardiología*, vol. 24, no. 2, pp. 110–116, Mar. 2017, doi: 10.1016/j.rccar.2016.06.013.
[25] O. Múñoz *et al.*, "Clinical practice guidelines for the prevention, early detection, diagnosis, treatment and follow-up of dyslipidemias: Pharmacological treatment with statins (in Spain)," *Revista Colombiana de Cardiología*, vol. 22, no. 1, pp. 14–21, Jan. 2015, doi: 10.1016/j.rccar.2015.02.001.
[26] O. Múñoz *et al.*, "Clinical practice guidelines for the treatment and follow-up of dyslipidemias in the population over 18 years of age (in Spain)," *Revista Colombiana de Cardiología*, vol. 22, no. 1, pp. 14–21, Jan. 2015, doi: 10.1016/j.rccar.2015.02.001.
[27] T. Tjur, "Coefficients of determination in logistic regression models—a new proposal: the coefficient of discrimination," *The American Statistician*, vol. 63, no. 4, pp. 366–372, Nov. 2009, doi: 10.1198/tast.2009.08210.
[28] Y. Zhang, S.-L. Guo, L.-N. Han, and T.-L. Li, "Application and exploration of big data mining in clinical medicine," *Chinese Medical Journal*, vol. 129, no. 6, pp. 731–738, Mar. 2016, doi: 10.4103/0366-6999.178019.
[29] S. Raschka and V. Mirjalili, *Python machine learning : machine learning and deep learning with python*. Packt Publishing, 2019.

BIOGRAPHIES OF AUTHORS

Astrid Lorena Urbano-Cano received the bachelor's degree in biology from Cauca University, Popayán, in 2003 and the M.S. and Ph.D. degrees in Biomedical Science from Antioquia University and Valle University, Colombia, in 2011 and 2018, respectively. Currently, she is an assistant Professor at the Department of Biology, Cauca University. The research interests include genetic, statistics, cardiovascular disease and teaches several courses such as genetics, molecular epidemiology, and investigation. She worked as researcher on projects by the Ministry of Science, Tech and Innovation, Republic of Colombia. She can be contacted at email: alurbano@unicauca.edu.co.

Diana Jimena López-Mesa received bachelor's degree in industrial automatic engineering and master's degree in automatic engineering, from Electronic, Instrumentation and Control Department, Universidad del Cauca, Popayán, Colombia, in 2009 and 2014, respectively. She is Ph.D. student in electronic sciencies in Universidad del Cauca, and has taught several courses such as electric machines, industrial process control, nonlinear control, descriptive and inferential statistics, mathematical logic, software for industrial applications. His main research focus is on electric drives, image processing, predictive models, and machine learning techniques. E-mail: djlopez@unicauca.edu.co.

Rosa Elvira Alvarez-Rosero received the bachelor's degree in biology from Cauca University, Popayán, in 1996 and the M.S. and Ph.D. degrees in biomedical science and environmental sciences from Valle University and Cauca University, Colombia, in 2013 and 2022, respectively. Currently, she is an Associate Professor at the Department of Physiological Sciences, Cauca University. The research interests include genetic, enviromental, cardiovascular disease and teaches several courses such as genetics and biology. She worked as researcher on projects by the Ministry of Science, Tech and Innovation, Republic of Colombia. She can be contacted at email: ralvares@unicauca.edu.co.

Yeison Alberto Garces-Gómez received bachelor's degree in electronic engineering, and master's degrees and Ph.D. in engineering from Electrical, Electronic and Computer Engineering Department, Universidad Nacional de Colombia, Manizales, Colombia, in 2009, 2011 and 2015, respectively. He is a full Professor at the Academic Unit for Training in Natural Sciences and Mathematics, Universidad Católica de Manizales, and teaches several courses such as experimental design, statistics and physics. His main research focus is on applied technologies, embedded systems, power electronics, power quality, but also many other areas of electronics, signal processing and didactics. He published more than 30 scientific and research publications, among them more than 10 journal papers. He worked as principal researcher on commercial projects and projects by the Ministry of Science, Tech and Innovation, Republic of Colombia. He can be contacted at email: ygarces@ucm.edu.co.