



Article

Predicting Cardiovascular Risk in Athletes: Resampling Improves Classification Performance

Davide Barbieri ^{1,†} , Nitesh Chawla ², Luciana Zaccagni ^{1,3,*} , Tonći Grgurinović ⁴, Jelena Šarac ⁵, Miran Čoklo ⁵ and Saša Missoni ^{6,7}

¹ Department of Biomedical and Specialty Surgical Sciences, Faculty of Medicine, Pharmacy and Prevention, University of Ferrara, 44121 Ferrara, Italy; davide.barbieri@unife.it

² Interdisciplinary Center for Network Science and Applications, University of Notre Dame, Notre Dame, IN 46556, USA; nchawla@nd.edu

³ Biomedical Sport Studies Center, University of Ferrara, 44123 Ferrara, Italy

⁴ Polyclinic for Occupational Health and Sports of Zagreb Sports Association with Laboratory of Medical Biochemistry, 10000 Zagreb, Croatia; toncimd@gmail.com

⁵ Centre for Applied Bioanthropology, Institute for Anthropological Research, 10000 Zagreb, Croatia; jelena.sarac@inantro.hr (J.Š.); miran.coklo@inantro.hr (M.Č.)

⁶ Institute for Anthropological Research, 10000 Zagreb, Croatia; sasa.missoni@inantro.hr

⁷ School of Medicine, Josip Juraj Strossmayer University of Osijek, 31000 Osijek, Croatia

* Correspondence: luciana.zaccagni@unife.it

† From 1 November 2020 the name of Department will be Neuroscience and Rehabilitation.

Received: 21 August 2020; Accepted: 25 October 2020; Published: 28 October 2020



Abstract: Cardiovascular diseases are the main cause of death worldwide. The aim of the present study is to verify the performances of a data mining methodology in the evaluation of cardiovascular risk in athletes, and whether the results may be used to support clinical decision making. Anthropometric (height and weight), demographic (age and sex) and biomedical (blood pressure and pulse rate) data of 26,002 athletes were collected in 2012 during routine sport medical examinations, which included electrocardiography at rest. Subjects were involved in competitive sport practice, for which medical clearance was needed. Outcomes were negative for the largest majority, as expected in an active population. Resampling was applied to balance positive/negative class ratio. A decision tree and logistic regression were used to classify individuals as either at risk or not. The receiver operating characteristic curve was used to assess classification performances. Data mining and resampling improved cardiovascular risk assessment in terms of increased area under the curve. The proposed methodology can be effectively applied to biomedical data in order to optimize clinical decision making, and—at the same time—minimize the amount of unnecessary examinations.

Keywords: medical diagnostic; decision tree; logistic regression; machine learning

1. Introduction

Cardiovascular diseases (CVDs) are reportedly the major cause of death worldwide, taking an estimated 17.9 million lives each year, according to the World Health Organization [1]. Obesity, smoking, physical inactivity and high blood pressure are among the most important risk factors [2]. Even if they may be mitigated by consistent sport practice [3,4], CVDs can still be considered an actual danger for individuals engaged in competitions or strenuous training [5–8] because of intense and repeated efforts. Therefore, athletes are routinely monitored by sport physicians, who collect some biomedical and personal data and screen them by means of electrocardiography (ECG).

According to the outcome of the ECG examination, individuals are diagnosed as either at risk (positive, P) or not (negative, N). Subjects at risk are denied medical clearance for sport practice

and eventually undergo further examination. The two classes are usually imbalanced, since the N class contains the majority of individuals, while the more interesting P class is under-represented. In general, a missing P (false negative, FN) may have a very high cost—in some cases the loss of a human life—while a false alarm (false positive, FP) usually has the cost of some further clinical investigations and temporary suspension of sport activities.

Classification is a *machine learning* technique which can be applied to predict categorical binary values, like P or N, and for such reason it may be of great value in the medical field, and in diagnostics in particular. Machine learning, a branch of artificial intelligence, consists in the application of computer algorithms in order to (semi) automatically extract knowledge from collected data. When classification is applied to large datasets, we usually speak of *data mining*, which is defined as “the process of discovering patterns in data (...) The patterns discovered must be meaningful in that they lead to some advantage, usually an economic advantage” [9]. As the amount of collected data has increased, researchers and physicians are interested in evaluating their diagnostic value by means of data mining, and eventually suggest that the observed variables may be changed or increased in order to support medical decisions. Several data mining methods have already been used as decision support systems for medical diagnosis [10]. These methods may be applied to large datasets in order to estimate health risk. Wong et al. [11] used Bayesian networks for early disease outbreak detection and obtained good performances on real data from an emergency department database, containing 7 years of medical data. The aim of the study was not diagnostic, since the data were collected from hospital patients who were actually ill, but rather epidemiological: to verify an incipient influenza outbreak (this approach could be adopted also in counter-terrorism, to detect a biological attack).

Campbell and Bennet [12] adopted a kernel-based method, which performed well on a medical dataset in identifying a rare disease. Still, the dataset size was limited and the proportion of interesting instances in the test set was very high (27 normal observations and 67 anomalies), compared to the prevalence of the disease in the general population.

Marinić et al. [13] adopted WEKA as a data mining tool. They applied a random forest classifier on a relatively small sample ($n = 102$) of psychiatric patients in order to diagnose Post Traumatic Stress Disorder (PTSD) and achieved significant results. Class distribution though was perfectly balanced (51 P and 51 N) and therefore different from that of the general population, where PTSD has a much lower prevalence. In addition, Fontaine et al. [14] explored data mining techniques in order to improve clinical evaluation of patients with neuropsychiatric disorders.

A data mining approach was proposed by Salam and McGrath [15] in dermatology. A multi-disease classifier improved medical diagnosis of skin disorders. Sacchi et al. [16] adopted a Naïve Bayes classification algorithm for the prediction of glaucoma. Having a small and imbalanced dataset, they applied both bootstrapping and resampling to train the model. Chan et al. [17] showed that machine learning classifiers outperformed traditional statistical approaches in the diagnosis of the same medical condition.

An increasing interest in the adoption of data mining for classification and prediction in cardiology was reported by Kadi et al. [18] in a recent systematic literature review. For example, Karaolis et al. [19] adopted a decision tree (DT) for the assessment of coronary risk.

Comparing different classification methods in medical statistics has been suggested by several authors [20–22] in order to assess the real advantages of one technique over the other. Still, there is a lack of studies applied to large datasets in domains where diseases have a low prevalence (like CVDs in sport medicine) but individuals may be at greater risk because of increased stress or pressure. Further, there is an on-going debate on the necessity of a sustainable and cost-effective health care [23], also by means of a more sensible use of medical tests [24].

The aims of this study were to assess the performances of a data mining method in the prediction of ECG outcomes in an imbalanced dataset, when a resampling technique is applied, and to verify whether the results may be used to support clinical decision making. The underlying hypothesis was

that, given a limited set of predictive biomedical variables, data mining could achieve good predictive accuracy if the proper algorithms were trained with a large amount of data.

2. Materials and Methods

2.1. Sample

A dataset including medical examinations of 26,002 athletes, both sexes, was collected at the Polyclinic for Occupational Health and Sports in Zagreb (Croatia) by medical staff in 2012. All individuals were involved in competitive sport practice, for which medical clearance was needed. The following data were collected for all subjects: sex, age, height, weight, resting pulse rate, diastolic and systolic pressure, and ECG at rest (P or N). The largest majority (91.2%) of outcomes was N, while a minority (8.8%) was P.

This study is the result of the collaborative research project “Health status and life quality of athletes”, involving the Institute for Anthropological Research in Zagreb, the Department of Biomedical Sciences and Surgical Specialties of the University of Ferrara, the Polyclinic for Occupational Health and Sports of Zagreb Sports Association with Laboratory of Medical Biochemistry in Zagreb, and the Interdisciplinary Center for Network Science and Applications of the University of Notre Dame. The research was approved by the Ethical Committee of the Institute for Anthropological Research in Zagreb (registration number: 1.14-1169/13).

2.2. Machine Learning Background

Two classification techniques were trained and tested in order to predict the class (P or N) of the athletes. DT was chosen because it allows us to describe the extracted knowledge (patterns) in a simple and intuitive way, which can be easily understood by domain experts, like medical doctors. It is commonly preferred when explanation (understanding) is as important as prediction (knowing). Further, DT is a well-established support tool in medical decision making [25–29].

Logistic regression (LR) is a technique commonly applied to medical datasets, half-way between classic statistics and machine learning. The main difference between regression and classification is that in the former the predicted variable is numeric, while in the latter it is categorical. Since in logistic regression, the predicted variable can have only two numeric values (1 or 0), it can be used for binary classification [30], where 1 stands for P and 0 for N.

The assessment of classification performance is a major issue in imbalanced datasets. Usually, the basic evaluation index is accuracy, the rate of correct guesses (i.e., true positives, TP, and true negatives, TN) on total instances $(TP + TN)/(P + N)$. It is an acceptable choice when class distribution is symmetric or close to it.

In case the distribution is imbalanced, accuracy can be misleading, unless a trivial solution is acceptable [31]. In fact, given a low prevalence, a high accuracy can be easily achieved classifying all instances as N, but it would imply missing all P individuals. Therefore, classification algorithms may have a high specificity or true negative rate ($TNR = TN/N$), as the majority class (N) is well represented, while sensitivity, or true positive rate ($TPR = TP/P$) may be significantly lower, as the minority class (P) is under-represented. Therefore, performance indexes other than accuracy should be taken into consideration.

A trade-off between TPR and TNR can be represented in receiver operating characteristic (ROC) space (for an introduction see [32]). Different cut-off values for the same classifier correspond to points in ROC space. The interpolation of such points draws a curve. The area under the ROC curve (AUC) is the accepted standard in the assessment of classification performances in imbalanced datasets [33], in particular in diagnostic systems [34–36].

Youden index $J = TPR + TNR - 1$ [37] is a summary measure corresponding to the distance between a point (corresponding to a cut-off value) on the ROC curve and the underlying 45° (random-guessing) line. J represents the probability of an informed decision. It has been used to assess the ability of

biomarkers to correctly classify healthy and non-healthy individuals when equal weight is given to both sensitivity and specificity [38,39].

Different resampling methods [40] can be applied in order to balance the class distribution and thus improve performances. Resampling can either undersample the majority class and/or oversample the minority class. Synthetic Minority Oversampling Technique (SMOTE) [41] has proved to be reliable in different domains, including the prediction of type 2 diabetes [42], SMOTE does not simply duplicate existing instances, which would easily lead to overfitting. It creates a new instance in feature (i.e., variable, in data mining jargon) space, between an existing instance and one, randomly-chosen, of its k nearest neighbors. Euclidean distance between two neighboring instances is calculated and then it is multiplied by a random number between 0 and 1. The distance is used to calculate the position in feature space of the new instance.

2.3. Data Cleaning and Resampling

The raw dataset was cleaned, removing outliers. Body weight and height were replaced by body mass index ($BMI = \text{weight}/\text{height}^2$), a common proxy for obesity and cardiovascular risk in the general population [43–47]. Only systolic pressure was used, since systolic and diastolic blood pressures are correlated, and according to recent findings, the former is a better predictor of risk [48,49].

SMOTE was applied to the minority class with 100% oversampling (P instances were doubled). Higher percentages were tried but they did not improve classification or led to overfitting. Then random undersampling was applied to the majority class with an even distribution spread, in order to balance the two classes, so that class ratio P/N was set to 1.

2.4. Data Mining and Statistical Analyses

The dataset was divided into training (66% of total instances) and test (remaining 34%). DT was first applied without resampling (1st run), then with resampling (2nd run, $k = 5$). In the latter case, only training set was resampled, while test set kept the same distribution as the original subset (which is supposed to be similar to the population, giving the high cardinality of the sample). This prevented performances from being artificially high. Pulse rate was the best predictor in both cases. Two thresholds were identified: low (L) and high (H). Risk was low in between, moderate for pulse rate $<L$ and high for pulse rate $>H$.

LR was applied twice. In the 1st run it was applied to all the collected variables directly. In the 2nd run, a discrete variable was created, with the following values: 0 (low risk) for pulse rate values between L and H, 1 (moderate risk) for values $<L$ and 2 (high risk) for values $>H$. This variable replaced pulse rate in a data-driven way, instead of using a-priori values from the literature. Cross-validation (10-fold) was used to assess the model's performances and test whether this statistical regression technique could improve the performances of a standard machine learning algorithm like DT.

TPR, TNR, J and AUC were used to assess the performances of the adopted method, which is described in Figure 1.

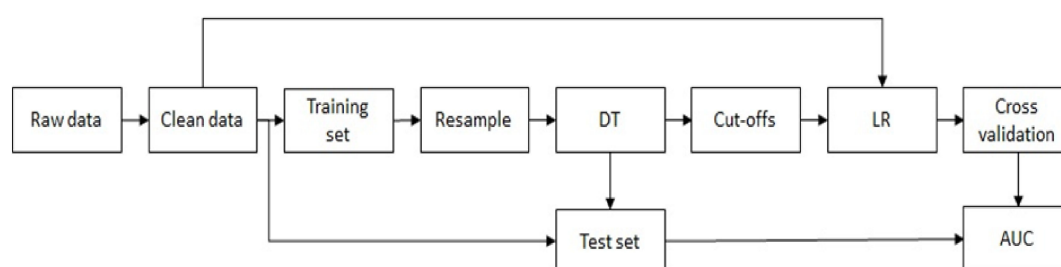


Figure 1. Data mining process. DT: decision tree; LR: logistic regression; AUC: area under curve.

MS Excel 2016 (Microsoft Corporation, Redmond, WA, USA) was used for data collection and data cleaning. WEKA 3.6 (University of Waikato, Waikato, New Zealand) was used for data mining.

Stata IC 13.1 (StataCorp LLC, College Station, TX, USA) was used to perform LR, cross validation and ROC analysis.

3. Results

Descriptive statistics by sex and age classes are shown in Table 1. Values are reported as mean and standard deviation, with the exception of ECG positives, which are reported as count and percentage.

Table 1. Descriptive statistics by sex and age classes.

| Variables | 6–10 Years | 11–14 Years | 15–18 Years | ≥19 Years | Total |
|----------------------------|-----------------|-----------------|-----------------|-----------------|-------------------|
| Females | <i>n</i> = 1372 | <i>n</i> = 1884 | <i>n</i> = 970 | <i>n</i> = 757 | <i>n</i> = 4983 |
| Weight (kg) | 34.2 ± 8.8 | 51.7 ± 10.8 | 61.8 ± 8.8 | 65.1 ± 10.5 | 50.9 ± 15.1 |
| Height (cm) | 138.3 ± 10.0 | 160.6 ± 8.7 | 168.5 ± 7.0 | 168.9 ± 7.52 | 157.3 ± 14.9 |
| BMI (kg/m ²) | 17.7 ± 2.8 | 19.9 ± 3.1 | 21.8 ± 2.6 | 22.8 ± 3.2 | 20.1 ± 3.5 |
| Pulse rate (<i>n</i>) | 83.1 ± 13.1 | 77.1 ± 13.0 | 68.3 ± 11.1 | 65.4 ± 11.7 | 75.2 ± 14.1 |
| Systolic pressure (mm Hg) | 97.3 ± 9.7 | 105.8 ± 9.9 | 108.7 ± 10.4 | 113.2 ± 11.8 | 105.2 ± 11.6 |
| Diastolic pressure (mm Hg) | 62.8 ± 7.47 | 66.4 ± 7.8 | 68.6 ± 7.8 | 73.3 ± 8.4 | 66.9 ± 8.6 |
| ECG Ps (<i>n</i> (%)) | 140 (10.2) | 160 (8.5) | 63 (6.5) | 65 (8.6) | 428 (8.6) |
| Males | <i>n</i> = 4787 | <i>n</i> = 5776 | <i>n</i> = 4253 | <i>n</i> = 6203 | <i>n</i> = 21,019 |
| Weight (kg) | 33.6 ± 8.6 | 52.4 ± 13.4 | 71.4 ± 11.7 | 83.9 ± 12.7 | 61.3 ± 22.6 |
| Height (cm) | 136.9 ± 9.2 | 161.4 ± 11.6 | 178.6 ± 7.6 | 180.0 ± 7.3 | 164.8 ± 19.2 |
| BMI (kg/m ²) | 17.7 ± 2.8 | 19.9 ± 3.3 | 22.3 ± 3.0 | 25.9 ± 3.4 | 21.6 ± 4.4 |
| Pulse rate (<i>n</i>) | 79.0 ± 12.6 | 73.1 ± 12.6 | 66.8 ± 12.4 | 63.8 ± 11.9 | 70.4 ± 13.7 |
| Systolic pressure (mm Hg) | 97.1 ± 9.1 | 106.8 ± 11.1 | 118.2 ± 10.9 | 126.4 ± 11.8 | 112.7 ± 15.6 |
| Diastolic pressure (mm Hg) | 61.5 ± 7.8 | 65.2 ± 7.8 | 69.6 ± 8.2 | 77.9 ± 9.4 | 69.0 ± 10.5 |
| ECG Ps (<i>n</i> (%)) | 379 (7.9) | 405 (7.0) | 397 (9.3) | 699 (11.3) | 1879 (8.9) |

BMI: body mass index; ECG Ps: electrocardiography positives.

Classification performances of DT and LR are shown in Table 2.

Table 2. Algorithm classification performances.

| Algorithm | TPR | TNR | J | AUC |
|--------------|------|------|------|------|
| DT (1st run) | 0.29 | 0.97 | 0.26 | 0.68 |
| LR (1st run) | 0 | 1 | 0.00 | 0.56 |
| DT (2nd run) | 0.68 | 0.82 | 0.50 | 0.76 |
| LR (2nd run) | 0.65 | 0.82 | 0.47 | 0.78 |

TPR: True positive rate; TNR: True negative rate; J: Youden index; AUC: Area under the ROC curve; DT: Decision tree; LR: Logistic regression.

DT performances improved considerably by means of resampling, particularly in terms of greater sensitivity, from 0.29 to 0.68.

LR was highly significant in both runs ($p < 0.001$). Still, in the first run, with the default cut-off = 0.5, performance was so low (AUC = 0.56) that no further attempt with other cut-off values was made. DT performed better, even without resampling.

After a discrete variable was introduced in place of pulse rate, LR increased sensitivity (with cut-off = 0.12) from 0 to 0.65, without considerably diminishing specificity. This result was achieved because risk was high in both the lower and upper range of the continuous variable, and not monotonously increasing. It became evident after DT was applied, since it classified as P instances with either low or high pulse rate values. L and H were found inductively in the data (L = 60 bpm, H = 99 bpm), but they were close to those which can be found in domain-specific literature.

4. Discussion

In the present study, we tested the effectiveness and accuracy of a data mining methodology in the prediction of cardiovascular risk in athletes. The application of resampling and two classification

techniques on an imbalanced dataset was evaluated. Resampling improved classification by means of DT. LR has proved to be not an accurate diagnostic tool if applied directly to continuous variables, since in the medical domain high risk is often associated with feature values which are either too low or too high. This is particularly true in the assessment of cardiovascular risk, since pulse rate and blood pressure have upper and lower thresholds.

Feature cut-off values, which were acquired inductively from the collected data, can be immediately converted into a small set of simple decision rules to assist the diagnostic process. Additionally, they can be used to create categorical or discrete variables for LR, in order to improve its performances, without the need for a-priori assumptions. The chosen predictors can be easily measured by a family practitioner at almost no cost, which would make cardiovascular risk evaluation very efficient.

Classification performances were not assessed by means of accuracy because of the asymmetric distribution of the two classes. AUC and J have been suggested, because they represent a trade-off between sensitivity and specificity, in case they are both given the same importance. Even if this assumption can be reasonable from a statistical point of view, in medical diagnostics it poses a serious ethical concern. In fact, a FP comes at a cost of some unnecessary examinations and temporary suspension of sport activities, while a FN may imply the loss of a human life, in case a serious condition is present. Since a FN has a higher cost than a FP, a cost matrix may be adopted, where correct predictions (TP and TN) have no cost, and errors (FN and FP) have different weights [50]. Still, in the medical domain, it is not always possible to compare the cost of errors. Since there is no general agreement on the amount of resources which can be allocated to reduce the risk related to FN, it is not easy to give the proper weights to the two kinds of errors.

Even if resampling improved DT performances, in terms of both TPR and J, sensitivity remained lower than specificity, implying a high false negative rate. This may not be acceptable in a medical domain where a high risk may imply the loss of a human life. Increased oversampling may lead to overfitting, even if a minimum TPR may be required or imposed by medical authorities. Therefore, LR cut-off values may be lowered to improve sensitivity, even if it implies a large reduction in both specificity and J. Nonetheless, it is important to note that DT and LR performances were similar, and therefore DT alone can be adopted as a predictive algorithm, with the advantage over LR of improved understanding. DT confirmed to be a good means to capture and represent domain-specific (medical, in this case) knowledge, in an intuitive and easy-to-understand way, something which cannot be achieved by means of LR. These facts may even question the application to medical datasets of statistical techniques instead of machine learning, when a binary classification is required, as in diagnostics.

The main limitation of this study lies in the lack of comparison with a standard control method (beside ECG, which is used as a gold-standard) for cardiovascular risk assessment. In order to overcome this limitation, cardiologists and general practitioners as well should be involved in a future study from the design phase.

5. Conclusions

This study tested the effectiveness of a simple resampling technique in improving the assessment of cardiovascular risk, when data are imbalanced, as it is often the case in real-world situations. SMOTE improved the performances of DT in terms of greater AUC and sensitivity. In addition, DT produced actionable knowledge, which can be applied in the prediction of CVDs, diminishing the need for assumptions.

Further research is required to test the performances of the proposed approach in the general (i.e., non-athletic and older) population, to determine an acceptable and agreed classification performance index—which will highlight the best trade-off between medical risk and sustainable welfare—and to verify whether this data mining methodology can be applied to improve diagnosis and optimize healthcare policies, thanks to a reduction of unnecessary examinations.

Author Contributions: Conceptualization, D.B., N.C., and L.Z.; methodology, D.B., N.C., L.Z., and S.M.; formal analysis, D.B. and L.Z.; investigation, T.G.; data curation, D.B. and T.G.; writing—original draft preparation D.B., N.C., L.Z., T.G., J.Š., M.Č., and S.M.; writing—review and editing D.B., N.C., L.Z., T.G., J.Š., M.Č., and S.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors would like to thank the Polyclinic for Occupational Health and Sports of Zagreb Sports Association with Laboratory of Medical Biochemistry in Zagreb for providing the data. The authors would also like to acknowledge their late colleague Joško Sindik, from the Institute for Anthropological Research in Zagreb, who contributed substantially to the content of this paper as well as the whole project.

Conflicts of Interest: The authors declare no conflict of interest.

References

- World Health Organization. Cardiovascular Diseases. Available online: https://www.who.int/health-topics/cardiovascular-diseases/#tab=tab_1 (accessed on 20 September 2020).
- Mendis, S.; Puska, P.; Norrving, B. (Eds.) *Global Atlas on Cardiovascular Disease Prevention and Control*; World Health Organization: Geneva, Switzerland, 2011.
- Güçlü, M. Comparing Women Doing Regular Exercise with Sedentary Women in Terms of Certain Blood Parameters, Leptin Level and Body Fat Percentage. *Coll. Antropol.* **2014**, *38*, 453–458.
- Wronka, I.; Suliga, E.; Pawlińska-Chmara, R. Evaluation of Lifestyle of Underweight, Normal Weight and Overweight Young Women. *Coll. Antropol.* **2013**, *37*, 359–365.
- Duraković, Z.; Duraković, M.M.; Skavić, J. Arrhythmogenic right ventricular dysplasia and sudden cardiac death in Croatians' young athletes in 25 years. *Coll. Antropol.* **2011**, *35*, 793–796.
- Duraković, Z.; Duraković, M.M.; Skavić, J. Hypertrophic cardiomyopathy and sudden cardiac death due to physical exercise in Croatia in a 27-year period. *Coll. Antropol.* **2011**, *35*, 1051–1054. [[PubMed](#)]
- Duraković, Z.; Misigoj Duraković, M.; Skavić, J.; Tomljenović, A. Myopericarditis and sudden cardiac death due to physical exercise in male athletes. *Coll. Antropol.* **2008**, *32*, 399–401. [[PubMed](#)]
- Chatard, J.C.; Mujika, I.; Goiriena, J.J.; Carré, F. Screening young athletes for prevention of sudden cardiac death: Practical recommendations for sports physicians. *Scand. J. Med. Sci. Sports* **2016**, *26*, 362–374. [[CrossRef](#)] [[PubMed](#)]
- Witten, I.H.; Frank, E. *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed.; Morgan Kaufmann Publishers (Elsevier): San Francisco, CA, USA, 2005.
- Bellazzi, R.; Zupan, B. Predictive data mining in clinical medicine: Current issues and guidelines. *Int. J. Med. Inform.* **2008**, *77*, 81–97. [[CrossRef](#)] [[PubMed](#)]
- Wong, W.K.; Moore, A.; Cooper, G.; Wagner, M. Bayesian Network Anomaly Pattern Detection for Disease Outbreaks. In Proceedings of the 20th International Conference on Machine Learning, Washington, DC, USA, 21–24 August 2003; Fawcett, T., Mishra, N., Eds.; The AAAI Press: Manlo Park, CA, USA, 2003.
- Campbell, C.; Bennett, K. A linear programming approach to novelty detection. In Proceedings of the Conference on Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 3–8 December 2001; MIT Press: Cambridge, MA, USA, 2001; Volume 14.
- Marinić, I.; Supek, F.; Kovačić, Z.; Rukavina, L.; Jendričko, T.; Kozarić-Kovačić, D. Posttraumatic Stress Disorder: Diagnostic Data Analysis by Data Mining Methodology. *Croat. Med. J.* **2007**, *48*, 185–197. [[PubMed](#)]
- Fontaine, J.F.; Priller, J.; Spruth, E.; Perez-Iratxeta, C.; Andrade-Navarro, M.A. Assessment of curated phenotype mining in neuro, psychiatric disorder literature. *Methods* **2015**, *74*, 90–96. [[CrossRef](#)]
- Salam, A.; McGrath, J.A. Diagnosis by numbers: Defining skin disease pathogenesis through collated gene signatures. *J. Investig. Dermatol.* **2015**, *135*, 17–19. [[CrossRef](#)]
- Sacchi, L.; Tucker, A.; Counsell, S.; Garway-Heath, D.; Swift, S. Improving predictive models of glaucoma severity by incorporating quality indicators. *Artif. Intell. Med.* **2014**, *6*, 103–112. [[CrossRef](#)] [[PubMed](#)]
- Chan, K.; Lee, T.W.; Sample, P.A.; Goldbaum, M.H.; Weinreb, R.N.; Sejnowski, T.J. Comparison of machine learning and traditional classifiers in glaucoma diagnosis. *IEEE Trans. Biomed. Eng.* **2002**, *49*, 963–974. [[CrossRef](#)] [[PubMed](#)]
- Kadi, I.; Idri, A.; Fernandez-Aleman, J.L. Knowledge discovery in cardiology: A systematic literature review. *Int. J. Med. Inform.* **2017**, *97*, 12–32. [[CrossRef](#)] [[PubMed](#)]

19. Karaolis, M.A.; Moutiris, J.A.; Hadjipanayi, D.; Pattichis, C.S. Assessment of the risk factors of coronary heart events based on data mining with decision trees. *IEEE Trans. Inf. Technol. Biomed.* **2010**, *14*, 559–566. [[CrossRef](#)]
20. Schwarzer, G.; Vach, W.; Schumacher, M. On the misuses of artificial neural networks for prognostic and diagnostic classification in oncology. *Stat. Med.* **2000**, *19*, 541–561. [[CrossRef](#)]
21. Zhang, S.; Tjortjis, C.; Zeng, X.; Qiao, H.; Buchan, I.; Keane, J. Comparing data mining methods with logistic regression in childhood obesity prediction. *Inf. Syst. Front.* **2009**, *11*, 449–460. [[CrossRef](#)]
22. Maroco, J.; Silva, D.; Rodrigues, A.; Guerreiro, M.; Santana, I.; de Mendonça, A. Data mining methods in the prediction of Dementia: A real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests. *BMC Res. Notes* **2011**, *4*, 299. [[CrossRef](#)]
23. Hood, V.L.; Weinberger, S.E. High value, cost-conscious care: An international imperative. *Eur. J. Intern. Med.* **2018**, *23*, 495–498. [[CrossRef](#)]
24. Qaseem, A.; Alguire, P.; Dallas, P.; Feinberg, L.E.; Fitzgerald, F.T.; Horwitch, C.; Humphrey, L.; LeBlond, R.; Moyer, D.; Wiese, J.G.; et al. Appropriate use of screening and diagnostic tests to foster high-value, cost-conscious care. *Ann. Intern. Med.* **2012**, *156*, 147–149. [[CrossRef](#)]
25. Murphy, C.K. Identifying diagnostic errors with induced decision trees. *Med. Decis. Mak.* **2001**, *21*, 368–375. [[CrossRef](#)]
26. Tanner, L.; Schreiber, M.; Low, J.G.; Ong, A.; Tolfvenstam, T.; Lai, Y.L.; Ng, L.C.; Leo, Y.S.; Thi Puong, L.; Vasudevan, S.G.; et al. Decision tree algorithms predict the diagnosis and outcome of dengue fever in the early phase of illness. *PLoS Negl. Trop. Dis.* **2008**, *2*, e196. [[CrossRef](#)]
27. Azar, A.T.; El-Metwally, S.M. Decision tree classifiers for automated medical diagnosis. *Neural Comput. Applic.* **2013**, *23*, 2387–2403. [[CrossRef](#)]
28. Christopher, J.J.; Nehemiah, H.K.; Kannan, A. A Swarm Optimization approach for clinical knowledge mining. *Comput. Methods Programs Biomed.* **2015**, *121*, 137–148. [[CrossRef](#)] [[PubMed](#)]
29. Gopinath, B.; Shanthi, N. Development of an Automated Medical Diagnosis System for Classifying Thyroid Tumor Cells using Multiple Classifier Fusion. *Technol. Cancer Res. Treat.* **2015**, *14*, 653–662. [[CrossRef](#)] [[PubMed](#)]
30. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning with Applications in R*; Springer: New York, NY, USA, 2013.
31. Provost, F.; Fawcett, T. Analysis and Visualization of Classifier Performance: Comparison under Imprecise Cost and Classifier Distribution. In Proceedings of the Third International Conference on Knowledge Discovery and Data Mining (KDD), Huntington Beach, CA, USA, 14–17 August 1997.
32. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874. [[CrossRef](#)]
33. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [[CrossRef](#)]
34. Swets, J.A. Measuring the accuracy of diagnostic systems. *Science* **1988**, *240*, 1285–1293. [[CrossRef](#)]
35. Ichikawa, D.; Saito, T.; Oyama, H. Impact of predicting health-guidance candidates using massive health check-up data: A data-driven analysis. *Int. J. Med. Inform.* **2017**, *106*, 32–36. [[CrossRef](#)]
36. Shimoda, A.; Ichikawa, D.; Oyama, H. Prediction models to identify individuals at risk of metabolic syndrome who are unlikely to participate in a health intervention program. *Int. J. Med. Inform.* **2018**, *111*, 90–99. [[CrossRef](#)]
37. Youden, W.J. Index for rating diagnostic tests. *Cancer* **1950**, *3*, 32–35. [[CrossRef](#)]
38. Ruopp, M.D.; Perkins, N.J.; Whitcomb, B.W.; Schisterman, E.F. Youden Index and optimal cut-point estimated from observations affected by a lower limit of detection. *Biom. J.* **2008**, *50*, 419–430. [[CrossRef](#)] [[PubMed](#)]
39. Perkins, N.J.; Schisterman, E.F. The Youden Index and the optimal cut-point corrected for measurement error. *Biom. J.* **2005**, *47*, 428–441. [[CrossRef](#)]
40. Lee, P.H. Resampling methods improve the predictive power of modeling in class-imbalanced datasets. *Int J. Environ. Res. Public Health* **2014**, *11*, 9776–9789. [[CrossRef](#)] [[PubMed](#)]
41. Chawla, N.V. Data Mining for Imbalanced Datasets: An Overview. In *Data Mining and Knowledge Discovery Handbook*; Maimon, O., Rokach, L., Eds.; Springer: New York, NY, USA, 2005; pp. 853–867.

42. Ramezankhani, A.; Pournik, O.; Shahrabi, J.; Azizi, F.; Hadaegh, F.; Khalili, D. The Impact of Oversampling with SMOTE on the Performance of 3 Classifiers in Prediction of Type 2 Diabetes. *Med. Decis. Mak.* **2016**, *36*, 137–144. [[CrossRef](#)]
43. Flegal, K.M.; Shepherd, J.A.; Looker, A.C.; Graubard, B.I.; Borrud, L.G.; Ogden, C.L.; Harris, T.B.; Everhart, J.E.; Schenker, N. Comparisons of percentage body fat, body mass index, waist circumference, and waist-stature ratio in adults. *Am. J. Clin. Nutr.* **2009**, *89*, 500–508. [[CrossRef](#)] [[PubMed](#)]
44. Freedman, D.S.; Kahn, H.S.; Mei, Z.; Grummer-Strawn, L.M.; Dietz, W.H.; Srinivasan, S.R.; Berenson, G.S. Relation of body mass index and waist-to-height ratio to cardiovascular disease risk factors in children and adolescents: The Bogalusa Heart Study. *Am. J. Clin. Nutr.* **2007**, *86*, 33–40. [[CrossRef](#)]
45. Zaccagni, L.; Lunghi, B.; Barbieri, D.; Rinaldo, N.; Missoni, S.; Šaric, T.; Šarac, J.; Babic, V.; Rakovac, M.; Bernardi, F.; et al. Performance prediction models based on anthropometric, genetic and psychological traits of Croatian sprinters. *Biol. Sport* **2019**, *36*, 17–23. [[CrossRef](#)]
46. Lam, B.C.; Koh, G.C.; Chen, C.; Wong, M.T.; Fallows, S.J. Comparison of Body Mass Index (BMI), Body Adiposity Index (BAI), Waist Circumference (WC), Waist-To-Hip Ratio (WHR) and Waist-To-Height Ratio (WHtR) as predictors of cardiovascular disease risk factors in an adult population in Singapore. *PLoS ONE* **2015**, *10*, e0122985. [[CrossRef](#)]
47. Suchanek, P.; Kralova Lesna, I.; Mengerova, O.; Mrazkova, J.; Lanska, V.; Stavek, P. Which index best correlates with body fat mass: BAI, BMI, waist or WHR? *Neuro Endocrinol. Lett.* **2012**, *33*, 78–82.
48. Borghi, C.; Dormi, A.; L'Italien, G.; Lapuerta, P.; Franklin, S.S.; Collatina, S.; Gaddi, A. The relationship between systolic blood pressure and cardiovascular risk—results of the Brisighella Heart Study. *J. Clin. Hypertens. (Greenwich)* **2003**, *5*, 47–52. [[CrossRef](#)]
49. Strandberg, T.E.; Pitkala, K. What is the most important component of blood pressure: Systolic, diastolic or pulse pressure? *Curr. Opin. Nephrol. Hypertens.* **2003**, *12*, 293–297. [[CrossRef](#)] [[PubMed](#)]
50. Weng Cheng, G.; Poon, J. A New Evaluation Measure for Imbalanced Datasets. In Proceedings of the 7th Australasian Data Mining Conference (AusDM '08), Glenelg/Adelaide, SA, Australia, 27–28 November 2008; Roddick, J.F., Li, J., Christen, P., Kennedy, P.J., Eds.; Australian Computer Society, Inc.: Darlinghurst, NSW, Australia, 2008; Volume 87, pp. 27–32.

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).