

Towards transparent machine learning models using feature sensitivity algorithm

Ali A. Abaker ^{a,1}, Fakhreldeen A. Saeed ^{a,2}

^aDepartment of Accounting Information Systems, Faculty of Computer Science,
AL Neelain University, Khartoum, Sudan

¹misterali88@gmail.com*, ²Email: fakhry00@gmail.com

ABSTRACT

Despite advances in health care, diabetic ketoacidosis (DKA) remains a potentially serious risk for diabetes. Directing diabetes patients to the appropriate unit of care is very critical for both lives and healthcare resources. Missing data occurs in almost all machine learning models, especially in production. Missing data can reduce the predictive power and produce biased estimates of models. Estimating a missing value around a 50 percent probability may lead to a completely different decision. The objective of this paper was to introduce a feature sensitivity score using the proposed feature sensitivity algorithm. The data were electronic health records contained 644 records and 28 attributes. We designed a model using a random forest classifier that predicts the likelihood of a developing patient DKA at the time of admission. The model achieved an accuracy of 80 percent using five attributes; this new model has fewer features than any model mentioned in the literature review. Also, Feature sensitivity score (FSS) was introduced, which identifies within feature sensitivity; the proposed algorithm enables physicians to make transparent, and accurate decisions at the time of admission. This method can be applied to different diseases and datasets.

Keywords:

Data analysis
Machine learning
Medical informatics
Feature sensitivity
DKA prediction

I. Introduction

The medical community has a consensus that diabetic ketoacidosis (DKA) is highly widespread in diabetes patients. According to [1], the incidence of DKA in the diabetes population is about 15.6 percent. DKA is responsible for up to four percent mortality. Diabetic ketoacidosis could potentially complicate the diagnosis, management, and prognosis of diabetes [2]. DKA is a life-threatening disease[3][4]. DKA has many psychosocial challenges for diabetes that put them at risk for repeated hospitalizations[5]. Diabetic ketoacidosis is a state of insulin deficiency that could lead to glucose disorder and lipid metabolism. DKA could, on the one hand, obscure and often delay the diagnosis of other diabetes complications[6]. On the other hand, it leads to more severe glucose hyperglycemia and intravascular volume depletion, both of which might increase morbidity and mortality [7].

Missing values happen due to periodic lack of sampling or input errors. These missing values lead to problems when working with machine learning[8][9]. Many studies handled this issue by removing missing values, but this approach suffers from the potential loss of significant information to support decision making [10][11]. The second method is replacement, which uses mean (for numeric attributes) or mode (for nominal attributes). To reduce the influence of unique values, the median can also be used [12]. Handling missing values is significant because machine learning models' performance is affected by missing data. Besides, most algorithms do not work on data with missing values [13]. Most of the introduced methods for handling missing values are on the training data [14].

This study aimed to predict the likelihood of a diabetic patient to develop DKA and identify the key features, which lead to DKA complication. We applied feature selection technique to reduce the number of features from the original data. We introduced a method to measure and identify within attributes sensitivity. The dataset was collected from Al Sukari Hospital for building the machine-learning model. The data set contained 730 records and 29 attributes. This model is useful for predicting the likelihood that a patient has a DKA or not, and the critical points within feature domain values, which have a significant impact on the physicians' decision-making process. The rest



of this paper is organized into Sections two to seven as follows: related works, data collection and preparation, methodology, experiments, results and discussion, future work, and conclusion.

II. Method

A. Method

We used the feature importance with the tree-based classifier as a method for reducing the number of features. Then we created categories for features. The feature sensitivity score is calculated for each category. Fig. 2 showed the general steps for the sensitivity algorithm.

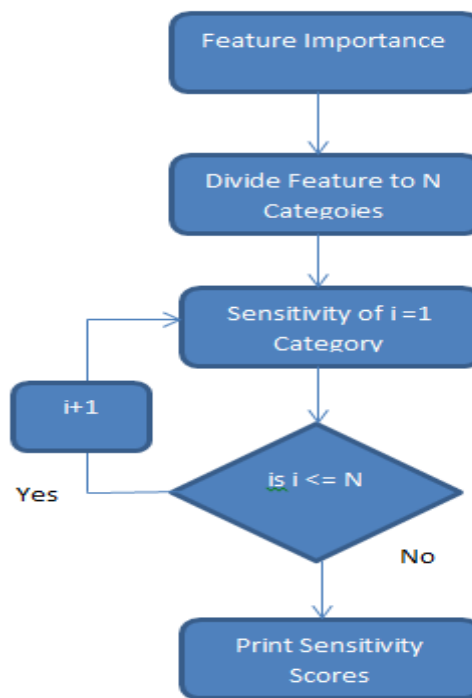


Fig. 1. General steps for the Sensitivity Algorithm.

Feature Sensitivity Algorithm in Fig. 2 was designed to calculate the sensitivity of each category of the features.

Algorithm 1: Feature Sensitivity Algorithm
Input: Data: Dataset; F: feature; M: model; Categories [2, n]: Array of start and end value of each category; n: Number of categories for a specific feature.
Output: sensitivity of this feature for each categories
 1: **repeat**
 2: Let $c=0$
 3: Let $D=$ select specific rows from Data where $F \geq \text{Categories}[0,c]$ and $F \leq \text{Categories}[1,c]$
 4: Let Total= number of rows in D
 5: Let TP= number of True Classes predicted by M
 6: Let $\text{FSensitivity} = \text{TP}/\text{Total}$
 7: Let $\text{CategoriesFS}[c] = \text{FSensitivity}$
 8: Let $c=c+1$
 9: **until** $c > n$
 10: **return** CategoriesFS

Fig. 2. Feature Sensitivity Algorithm

B. Related works

Brakel and Schrauwen [15] proposed two models for imputing missing values to train the model. The main idea of the imputation method is that if the value is not available for a particular instance, it could be estimated [9]. Fortes et al. [16] introduced a method to handle missing values with an

associated confidence and error parameter of the suggested value. Pelckmans et al. [17] proposed an approach for handling missing values taking into account the predicted outcome when missing values are taking place; this approach is based on the mean imputation. Rajawat et al. built a predictive analysis model using a hybrid machine learning technique. They filled attributes that had missing values with the median value of that attribute [18]. The previous studies depended on estimation when filling the missing values.

Efstathiou et al. [19] conducted a study to predict the mortality in diabetic ketoacidosis. The clinical and laboratory parameters values were accessed to predict the mortality in patients with diabetic ketoacidosis (DKA). They built a model using 20 attributes. The study suggested that simple clinical and laboratory parameters available in the first 24 hours from admission may contribute more objectively. However, they used many attributes, which may complicate the model and decrease performance. A. Deeb *et al*[20]. They proposed a model of care, to reduce the frequency of hospital admission of children and adults presented with DKA. The data was collected over four years of 158 admissions for DKA; they did not mention the attributes, which they used as predictors. S. Suwanto et al. [21]. Focused on building a model to predict DKA mortality, they used six attributes and 60 records. However, they used a limited number of observations. N. N. Siregar et al. [22] identified the predictors of 72-hour mortality in patients with DKA. They used 301 health records of adult patients; the results showed that 11 predictors are associated with DKA complications and mortality. The difference between the previous studies and this work that we used fewer attributes for predicting DKA. Furthermore, that reflected directly on the reduction of the training time.

C. Features Selection

The process of discovering and choosing the most significant features in a dataset is called feature selection. It is a vital step in the machine learning pipeline [23]. Irrelevant features that do not help a machine learning model: reduce training speed, reduce model interpretability, and, most importantly, reduce performance on the test set [24]. Though it looks simple, it is one of the most challenging processes in the work of designing a new machine learning model. Feature Selection and Data Cleaning should be the first step in creating the model.

The feature selection algorithms estimate feature importance based on the characteristics of the features, such as feature variance and relevance to the target variable. Selecting important features are part of a data pre-processing step and then train a model using the selected features [25]. Therefore, feature selection is uncorrelated to the training algorithm. It is a technique to present how the features in the model contribute to the model prediction. Several methods exist to get some insight into these black-box models. Feature importance gives a score for each feature of the data; the higher the score more important or relevant is the feature towards the output variable [26].

D. Data Collection and preparation

The dataset was collected from Alsukari Hospital. Ethical approval to use the data for research was obtained both from the Ministry of Health (MOH) and the hospital. The dataset contained 644 records and 28 attributes of diagnosed diabetes patients who were admitted to the hospital in the period from January 2018 to April 2019.

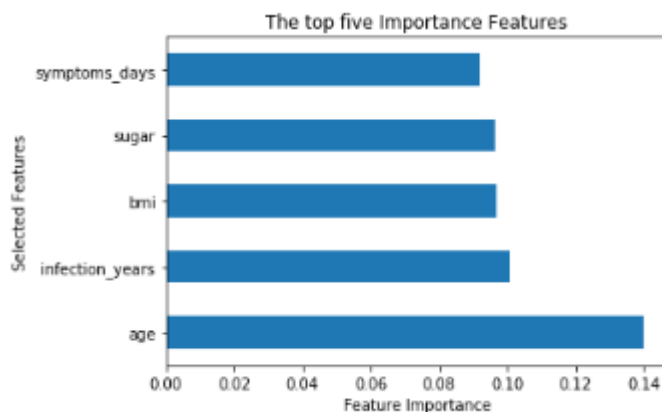


Fig. 3. The selected features and their importance.

The figure above shows the features and their contribution to the model. we can observe that age is the most important feature with about 14 percent, and we can notice that the rest of the four features have importance around 10 percent for each. Feature importance is the contribution of a feature in the model Adding more features means more information to the model. But, we have to balance between the number of features and the model performance. The proposed model was able to run using fewer attributes, and maintained a level of accuracy, by only 18 percent of the original features.

Table 1. The selected attributes used in the model and their scales.

No	Attribute	Type	Scale
	age	Continuous	1 – 97
	infection_years	Continuous	0 – 40
	BMI	Continuous	11 – 45
	sugar	Continuous	33 – 681
	Symptoms_days	Continuous	0 – 120
	class	Binary	0:No DKA, 1:Have DKA

E. Features Description

- **Age:** the age of the diabetes patient.
- **Infection_years:** The date since a patient was diagnosed with diabetes in (years).
- **BMI:** Body Mass Index (Table 2) is a measure of body fat based on height and weight that applies to adult men and women.

Table 2. Categorization of Body Mass Index developed by WHO panel[27].

Category	Quartiles Ranges
Underweight	<18.5
Normal	18.5–24.99
Overweight	25–29.99
Obese	30+

$$\text{BMI} = \text{weight (kg)} / [\text{height (m)}]^2 \quad (1)$$

- **Sugar:** Blood sugar, or glucose, is the main sugar found in blood. The level is measured by milligrams per deciliter (mg/dL) or using millimoles per liter (mmol/L).
- **Symptoms_days:** The date since the symptoms started appearing on the patient until arriving at the hospital.
- **Class:** The target variable.

III. Results and Discussion

A. Experiments

After Random Forest (RF) is a machine learning algorithm that is used widely for classification problems, RF is made of an ensemble of autonomous decision trees[21]. Each tree is learned with randomly selected samples and features[30]. We created a model to predict the test data. So, we use the training data to fit the model and testing data to test it. We used 70 percent of the data for training and 30 percent as a testing set. The oversampling method was implemented on the training set, for handling imbalance in the dataset. We also adjusted some hyperparameters of the algorithm, like `n_estimators`, `max_depth`, `random_state`, 40, 16, 18, respectively. The model achieved an accuracy of 80 percent on the testing data.

We used two features age and DKA (Fig. 4) for testing the efficiency of our proposed algorithm. The figures below visualize the results of the feature sensitivity algorithm.

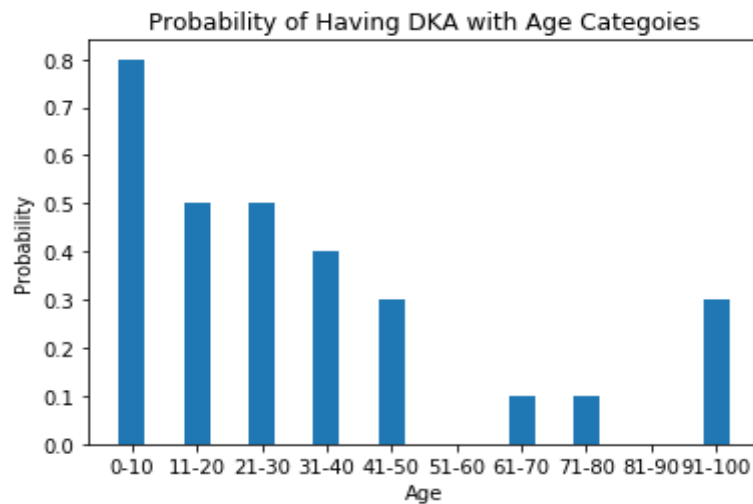


Fig. 4. Age categories with DKA probabilities.

From Fig. 4, it is clear that patients with age categories between 0-10 are most likely to have DKA with the probability of 80 percent, while patients between 11 to 30 are in the critical area. The feature sensitivity score (FSS) is about 50 percent; thus, the absence of the Age attribute value is critical for the decision making process as we are working in the healthcare field. However, patients who are older than 50 years have a lower probability of having DKA. Moreover, therefore are not in the critical area of decision making.

Table 3. Shows sensitivity type of each category within age feature.

Age Categories	Category Sensitivity	Feature Importance	Expected Probability	Sensitive
0-10	80%	(-) 0.14 %	66	No
11-20	50%		36	Yes
21-30	50%		36	Yes
31-40	40%		26	No
41-50	30%		16	No
51-60	0%		-14	No
61-70	10%		-4	No
71-80	10%		-4	No
81-90	0%		-14	No
91-100	30%		16	No

From Table 3, it is clear that 11-20 and 21-30 categories are sensitive in the age feature. 11-20 and 21-30 categories have 50% sensitivity, and considering the absence of their values, then we have 9.5% (feature importance) as error rate in the model outcome. So, we may have a different decision.

From Fig. 5, we can observe that patients whom their body mass index (BMI) category between 31-35 are most likely to have DKA with the probability of 50 percent. Missing BMI from this range is very sensitive for the decision making because this is the turning point for the threshold. However, patients who have BMI in categories 11-15, 21-25, and 26-30 have the lowest probability of having DKA.

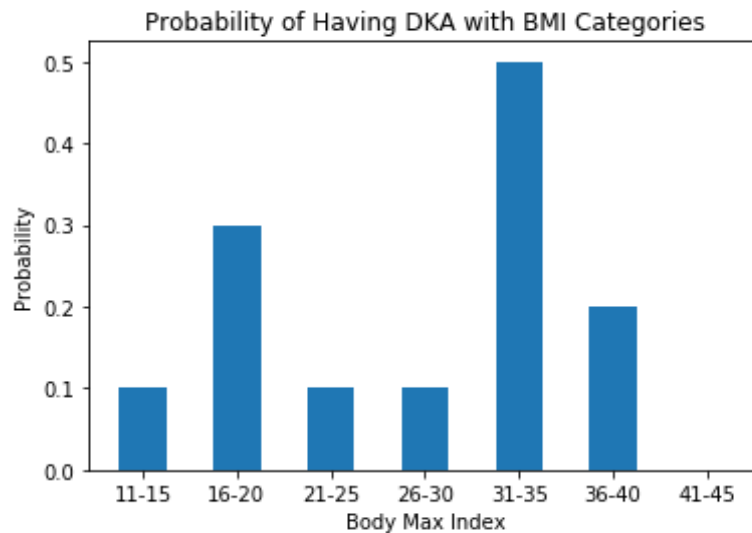


Fig. 5. BMI categories with DKA probabilities.

According to Table 4, we can notice that the 31-35 category is the only sensitive among body mass index(BMI) feature. 31-35 category has 50% sensitivity and considering the absence of it is value, then we have 9.5% (feature importance) as error rate in the model outcome. We may have a different decision.

Table 4. The sensitivity type of each category within the BMI feature.

BMI Categories	Category Sensitivity	Feature Importance	Expected Probability	Sensitive
11-15	10%	(-) 9.5 %	0.5	No
16-20	30%		20.5	No
21-25	10%		0.5	No
26-30	10%		0.5	No
31-35	50%		40.5	Yes
36-40	20%		10.5	No
41-45	0%		-9.5	No

B. Discussion

We built a machine learning model with an accuracy of 80 percent. The new model has fewer features compared to the previous models in the literature review. We introduced the feature sensitivity score (FSS) to identify within feature sensitivity using the proposed sensitivity algorithm. According to figure 3, the critical probability points within age feature are in 11-20, and 21-30 with a probability of 50 percent. Furthermore, from figure 4, Patients who are in the 31-35 body mass index (BMI) range are in the critical area because of the 50 percent probability.

The critical probability point means any under or overestimation of a missing value around 50 percent probability may lead to a completely different decision. The feature importance of the feature has to be taken into account. Using the proposed method would enable physicians to have transparency on model accuracy when making decisions.

IV. Conclusion

The findings of this study show that the proposed feature sensitivity algorithm approves the ability to determine within feature sensitivity. Feature sensitivity score will help doctors to make transparent and accurate decisions when missing any of the feature value. In the medical field, we need to make accurate decisions. Clinicians' decisions are one of the most important factors for guiding the cost and quality of medical care. Decisions help determine what prevention programs must be provided, what diagnoses to be made, what tests are requested, and what treatments are offered.

Acknowledgment

We would like to thank Alsukari Hospital for providing us with the dataset and extends thanks to the medical staff and the statistics department. Also, the Ministry of Health (MOH) for the ethical approval to use the dataset for this research.

References

- [1] A. Usman, S. A. Syed Sulaiman, A. H. Khan, and A. S. Adnan, "Profiles of diabetic ketoacidosis in multiethnic diabetic population of Malaysia," *Trop. J. Pharm. Res.*, vol. 14, no. 1, pp. 179–185, 2015, doi: 10.4314/tjpr.v14i1.25.
- [2] V. R. Balpande and R. D. Wajgi, "Prediction and severity estimation of diabetes using data mining technique," *IEEE Int. Conf. Innov. Mech. Ind. Appl. ICIMIA 2017 - Proc.*, no. Icimia, pp. 576–580, 2017, doi: 10.1109/ICIMIA.2017.7975526.
- [3] L. M. Duca, B. Wang, M. Rewers, and A. Rewers, "Diabetic ketoacidosis at diagnosis of type 1 diabetes predicts poor long-term glycemic control," *Diabetes Care*, vol. 40, no. 9, pp. 1249–1255, 2017, doi: 10.2337/dc17-0558.
- [4] P. Vellanki and G. E. Umpierrez, "Increasing hospitalizations for DKA: A need for prevention programs," *Diabetes Care*, vol. 41, no. 9, pp. 1839–1841, 2018, doi: 10.2337/dci18-0004.
- [5] M. A. Harris, D. V. Wagner, M. Heywood, D. Hoehn, H. Bahia, and K. Spiro, "Youth repeatedly hospitalized for dka: Proof of concept for novel interventions in children's healthcare (NICH)," *Diabetes Care*, vol. 37, no. 6, pp. 2013–2015, 2014, doi: 10.2337/dc13-2232.
- [6] K. Alexiadou and J. Doupis, "Management of diabetic foot ulcers," *Diabetes Ther.*, vol. 3, no. 1, pp. 1–15, 2012, doi: 10.1007/s13300-012-0004-9.
- [7] Y. Wang *et al.*, "Concurrent Diabetic Ketoacidosis in Hypertriglyceridemia-Induced Pancreatitis: How Does It Affect the Clinical Course and Severity Scores?," *Pancreas*, vol. 46, no. 10, pp. 1336–1340, 2017, doi: 10.1097/MPA.0000000000000937.
- [8] G. D. Betrie, R. Sadiq, S. Tesfamariam, and K. A. Morin, "Zur Problematik Von Unvollständigen Und Fehlenden Beschaffheitsdaten in Datenbanken Von Bergbaustandorten: Vergleich Von Drei Berechnungsmethoden," *Mine Water Environ.*, vol. 35, no. 1, pp. 3–9, 2016, doi: 10.1007/s10230-014-0322-4.
- [9] G. Chhabra, V. Vashisht, and J. Ranjan, "A classifier ensemble machine learning approach to improve efficiency for missing value imputation," *2018 Int. Conf. Comput. Power Commun. Technol. GUCON 2018*, pp. 23–27, 2019, doi: 10.1109/GUCON.2018.8674904.
- [10] H. He, Y. Cao, Y. Cao, and J. Wen, "Ensemble learning for wind profile prediction with missing values," *Neural Comput. Appl.*, vol. 22, no. 2, pp. 287–294, 2013, doi: 10.1007/s00521-011-0708-1.
- [11] Y. Liu and V. Gopalakrishnan, "An overview and evaluation of recent machine learning imputation methods using cardiac imaging data," *Data*, vol. 2, no. 1, 2017, doi: 10.3390/data2010008.
- [12] P. Liu, E. El-Darzi, L. Lei, C. Vasilakis, P. Chountas, and W. Huang, "An analysis of missing data treatment methods and their application to health care dataset," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 3584 LNAI, pp. 583–590, 2005, doi: 10.1007/11527503_69.
- [13] Doreswamy, I. Gad, and B. R. Manjunatha, "Performance evaluation of predictive models for missing data imputation in weather data," *2017 Int. Conf. Adv. Comput. Commun. Informatics, ICACCI 2017*, vol. 2017-Janua, pp. 1327–1334, 2017, doi: 10.1109/ICACCI.2017.8126025.
- [14] A. Kota Gopalakrishna, T. Özcelebi, A. Liotta, and J. J. Lukkien, "Treatment of missing data in intelligent lighting applications," *Proc. - IEEE 9th Int. Conf. Ubiquitous Intell. Comput. IEEE 9th Int. Conf. Auton. Trust. Comput. UIC-ATC 2012*, pp. 1–8, 2012, doi: 10.1109/UIC-ATC.2012.135.
- [15] P. Brakel and B. Schrauwen, "Energy-based temporal neural networks for imputing missing values," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7664 LNCS, no. PART 2, pp. 575–582, 2012, doi: 10.1007/978-3-642-34481-7_70.
- [16] I. Fortes, L. Mora-López, R. Morales, and F. Triguero, "Inductive learning models with missing values," *Math. Comput. Model.*, vol. 44, no. 9–10, pp. 790–806, 2006, doi: 10.1016/j.mcm.2006.02.013.
- [17] K. Pelckmans, J. De Brabanter, J. A. K. Suykens, and B. De Moor, "Handling missing values in support vector machine classifiers," *Neural Networks*, vol. 18, no. 5–6, pp. 684–692, 2005, doi: 10.1016/j.neunet.2005.06.025.
- [18] P. S. Rajawat, D. K. Gupta, S. S. Rathore, and A. Singh, "Predictive Analysis of Medical Data using a Hybrid Machine Learning Technique," *ICSCCC 2018 - 1st Int. Conf. Secur. Cyber Comput. Commun.*, pp. 228–233, 2018, doi: 10.1109/ICSCCC.2018.8703302.

- [19] S. P. Efstathiou *et al.*, "A mortality prediction model in diabetic ketoacidosis," *Clin. Endocrinol. (Oxf.)*, vol. 57, no. 5, pp. 595–601, 2002, doi: 10.1046/j.1365-2265.2002.01636.x.
- [20] A. Deeb *et al.*, "Implementation of a Diabetes Educator Care Model to Reduce Paediatric Admission for Diabetic Ketoacidosis," *J. Diabetes Res.*, vol. 2016, 2016, doi: 10.1155/2016/3917806.
- [21] S. Suwanto, B. Sutrisna, S. Waspadji, and H. T. Pohan, "Predictors of five days mortality in diabetic ketoacidosis patients: a prospective cohort study," *Acta Med. Indones.*, vol. 46, no. 1, pp. 18–23, 2014.
- [22] N. N. Siregar, P. Soewondo, I. Subekti, and M. Muhadi, "Seventy-two hour mortality prediction model in patients with diabetic ketoacidosis: A retrospective cohort study," *J. ASEAN Fed. Endocr. Soc.*, vol. 33, no. 2, pp. 124–129, 2018, doi: 10.15605/jafes.033.02.03.
- [23] A. Razmjoo, P. Xanthopoulos, and Q. P. Zheng, "Online feature importance ranking based on sensitivity analysis," *Expert Syst. Appl.*, vol. 85, pp. 397–406, 2017, doi: 10.1016/j.eswa.2017.05.016.
- [24] X. Deng, Y. Li, J. Weng, and J. Zhang, "Feature selection for text classification: A review," *Multimed. Tools Appl.*, vol. 78, no. 3, pp. 3797–3816, 2019, doi: 10.1007/s11042-018-6083-5.
- [25] D. Panda, R. Ray, A. A. Abdullah, and S. R. Dash, "Predictive Systems: Role of Feature Selection in Prediction of Heart Disease," *J. Phys. Conf. Ser.*, vol. 1372, no. 1, 2019, doi: 10.1088/1742-6596/1372/1/012074.
- [26] A. Zien, N. Krämer, S. Sonnenburg, and G. Rätsch, "The feature importance ranking measure," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 5782 LNAI, no. PART 2, pp. 694–709, 2009, doi: 10.1007/978-3-642-04174-7_45.
- [27] J. S. Markowitz, "Mortality and Its Risk Factors Among Professional Athletes. A Comparison Between Former NBA and NFL Players," pp. 39–49, 2018, doi: 10.1007/978-3-319-77203-5.
- [28] Z. Masetic and A. Subasi, "Congestive heart failure detection using random forest classifier," *Comput. Methods Programs Biomed.*, vol. 130, pp. 54–64, 2016, doi: 10.1016/j.cmpb.2016.03.020.
- [29] M. Monirul Kabir, M. Monirul Islam, and K. Murase, "A new wrapper feature selection approach using neural network," *Neurocomputing*, vol. 73, no. 16–18, pp. 3273–3283, 2010, doi: 10.1016/j.neucom.2010.04.003.
- [30] F. Baumann, F. Li, A. Ehlers, and B. Rosenhahn, "Thresholding a Random Forest classifier," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8888, pp. 95–106, 2014, doi: 10.1007/978-3-319-14364-4_10.