Bioengineering Research

Benefiting machine learning methods to detect Fraud in the validation of bank customers' cards

Parisa Ahmadi [1], Ahmad Yousefi [2*], and Farid Rezazadeh [1]

[1] Department of Computer, Faculty of Engineering, Aghigh University, Shahin shahr, Isfahan, Iran.

[2] Department of Computer, Naein Branch, Islamic Azad University, Naein, Isfahan, Iran

Correspondence to: Yousefi A. (E-mail: aaa.yousofi@gmail.com)

Abstract

The existence of money laundering and banking fraud is one of the major challenges of the banking system in any country. Crediting customers based on their track record and performance is a method to address the banking challenges. Classification methods can be used to validate customers, but these methods own high error. In this paper, machine learning methods are applied on banking data set to classify them and to reduce the error in customer validation. To this end, first the machine learning methods are trained and then tested using the banking data set. Experiments on the banking data set show that the accuracy of the proposed method for validating customers is less than 81.6%. So, the accuracy index of the random forest, decision tree, support vector machine, and multilayer artificial neural network are 80.50%, 80.05%, 80.93%, and 81.58%, respectively. The best performance is related to multilayer artificial neural network and, accordingly, the multilayer artificial neural network method can be used in detection of the validation of bank customers' cards.

Received: 4 May 2021, **Accepted**: 10 June 2021
DOI: 10.22034/jbr.2021.284617.1039
Keywords: Data Mining; Machine Learning; Bank Fraud

1. Introduction

Today, most financial services are provided through the virtual world and the Internet, and e-banking has developed significantly [1-2]. To increase their productivity, banks are creating virtual systems on the Internet, and this has made it easy for people to use its financial services without going to the bank [3-4]. Despite the development of web banking and finance in this context, some challenges make this advantage in a disadvantage, and that is the use of financial exchange platforms for illegal activities. Today, a large number of transactions are made within the banking system, that a large part of them are related to the Internet and with the help of Internet payment gateways [5-6]. Activities that use the financial platform for criminal activities try to hide their traces and legitimize their dirty money. Today, fraud [7] and money laundering [8] are two major challenges on the Internet, and these activities are carried out by criminals to cover up their illegal activities and in some cases individuals or companies to commit tax evasion. Consequently, they increase launder money and moving dirty money.

One of the important methods for detecting fraud and money laundering and analyzing banking transactions is to use knowledge discovery methods [9], which can be referred to the data mining and machine learning.

The latest pattern of abnormal transactions can be distinguished from normal transactions using these methods [10-11]. An important advantage of data mining methods is that by intelligently searching through large volumes of data and banking transactions to find patterns of fraud and money laundering [12]. Several methods have been proposed to detect money laundering and bank fraud based on learning, including deep learning [13], artificial neural network [14], support vector machine [15,16], decision tree [17], random forest [18] and regression [19] that these methods have been used for other fields such as biomedical [20], healthcare [21-22], rehabilitation [23-24].

The morphological classification and analysis for the validation of individuals and the diagnosis of fraud were investigated in a study [25]. In the mentioned study, a completed classification of scams was provided based on morphological analysis, feature list, and matrix analysis, and then the frameworks of cheat matrix and tree classification were presented and discussed. These frameworks were then used to classify and explain several different types of scams. To this end, first, the features related to fraud were compiled and then the channels of fraud and the basic features of fraud were introduced and modeled. Furthermore, several popular types of scams were identified using this mentioned method by benefiting the fraud classification framework in [25]. Besides, new types of fraud were detected using the proposed framework, for example, transaction fraud, automated fraud, concurrent fraud, and so on. The importance of classification is that it can be used to classify both existing and newly identified fraud that has not been previously reported. In another study [26], a framework for fraud identifying in credit cards was provided with a learning and cost-sensitive approach. Electronic payment systems also continue to assist businesses around the world, and credit cards are a means of payment in electronic payment systems. However, fraud due to the use of credit cards remains a major threat for financial institutions that there are various reports and statistics in this field. Several machine learning methods have been developed to reduce this prevailing threat in payment systems, in which group methods and cost-sensitive learning techniques play a key role. In their study in [26], a new framework was presented that combines the potential of learning techniques and a cost-sensitive learning pattern to detect fraud. The results obtained from the data classification illustrate that the cost-sensitive group classifier shows an excellent value in the AUC index, which indicates the accurate performance in detecting the fraud rate in the data set. Furthermore, the framework can effectively detect fraudulent transactions in different databases, and it is efficient regardless of the ratio of fraud cases compared to the performance of other classification methods.

In this paper, fraud detection in bank customers' cards is performed based on a machine learning algorithm. The obtained results illustrate the accuracy of random forest, decision tree, support vector machine, and multilayer artificial neural network are 80.50\%, 80.05\%, 80.93\%, and 81.58\%, respectively. Thus, the multilayer artificial neural network is efficient in fraud detection in the validation of bank customers' cards.

2. Material and Method

2.1. The Proposed Method

The proposed method for customer validation uses a classification technique based on a machine learning methods. The algorithms consist of random forest, decision tree, support vector machine, and multilayer artificial neural network. In these proposed methods, customers can be classified and validated based on the type of transactions.

The framework of validation of customers in the bank for the proposed method is performed based on their credit by performing fraudulent transactions or money laundering. The data and samples are divided into two categories of training and testing, and this ratio is equal to 70% to 30%, and the training data is used to train the artificial neural network. In the proposed method, several data set samples are considered as model inputs, and their classification is performed. Classification error or classification accuracy index can be used to evaluate the proposed method.

2.2. Data Set

One of the practical methods for analyzing different types of data sets is to use the Weka analysis tool to evaluate the data collection and the validation of bank customers. The database was downloaded from the database set at the UCI site [26]. A view of this data set in the Weka environment is shown in Fig. 2 that it has 30,000 records, and each record owns several features. An important advantage of using Weka tools is that the data set can be well analyzed, and basic learning methods can be implemented by it, and this process does not require programming. Implementation of data mining algorithms and methods such as multilayer artificial neural network, recursive artificial neural network, support vector machine, decision tree, and random forest on all types of data can be done by the Weka software. MATLAB software is used to analyze the proposed method, which requires programming, and it can be compared with similar methods in Weka or other studies to compare the proposed method.

In this data set, 30,000 samples of customers have been used, of which 23,364 were reputable customers and 6,636 of them were involved in money laundering and bank fraud and did not have credit in the banking system. In Fig. 3, some records with their values are shown in this dataset that 23 attributes are input, and the last feature is output. In this data set, the feature of 1 to 23 is the input, and the 24th attribute is of the output, which its value can be zero and one and represents the valid or invalid customer, respectively. 23 of features contain key information about customers, including account number, age, gender, level of education, current account amount, one-year transactions per year, and other features about customer payments. The output feature also provides both normal and abnormal modes for the customer, and therefore the problem can be considered as a two-class classification.

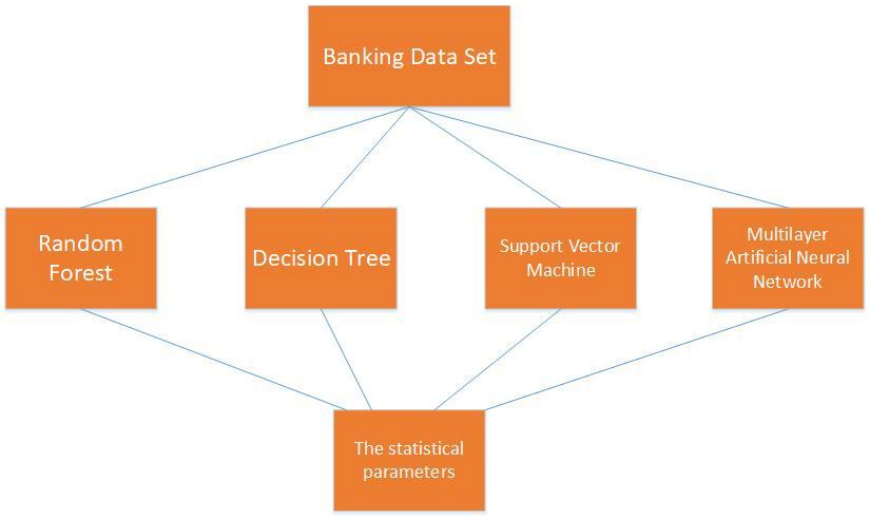

Figure 1. The proposed machine learning algorithms for customer validation

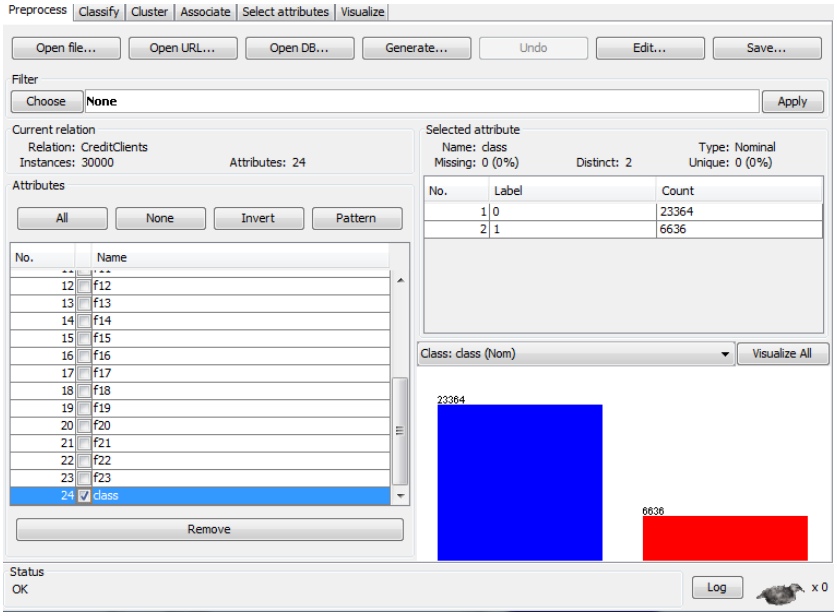

Figure 2. Banking customer validation database in Weka tools.

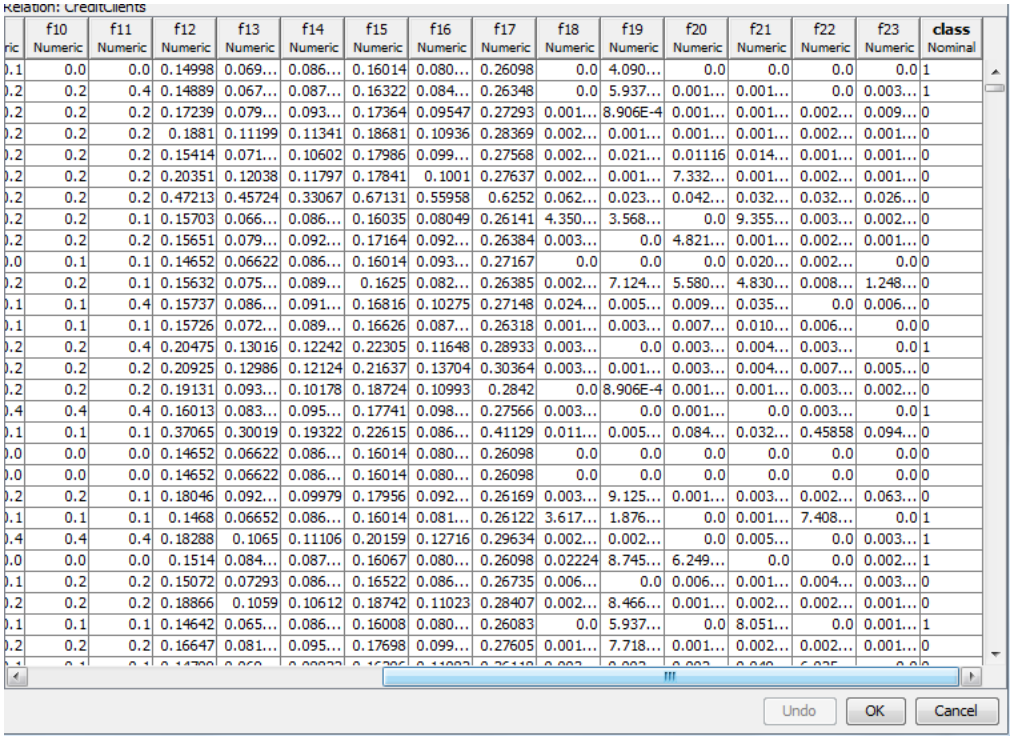

Figure 3. Several records for accreditation of bank customer.

2.3. Credit indicators of bank customers

Criteria such as error and accuracy of classification are used in most researches in the field of credit of bank customers. In this paper, these indicators such as error and accuracy of customer classification are used. Eq. 5 shows the mean absolute error (MAE) to evaluate the proposed method in analyzing customer behavior. Because there are only two outputs of zero and one in the output of this data set, the MAE is equal to the MSE index in Eq. 6 [27].

$$MAE = \frac{1}{n}\sum_{k=1}^{n}\left|\widetilde{O}_{\iota} - O_i\right| \quad (1)$$

$$MSE = \frac{1}{n}\sum_{k=1}^{n}\left(\widetilde{O}_{\iota} - O_i\right) \quad (1)$$

Where $\widetilde{O}_{\iota}$ and O_i are the actual and predicted numbers of a customer's class or transaction in terms of normal and abnormal types, also n is the number of customer evaluation samples. In addition to the error-index, rating systems of the bank customer credit such as accuracy, sensitivity, and precision can be evaluated that are presented in Eq. 7-9, respectively [28].

$$Accuracy = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \quad (1)$$

$$Sensitivity = \frac{T_p}{T_p + T_n} \quad (2)$$

$$Precision = \frac{T_p}{F_n + T_p} \quad (3)$$

The value of each of these indicators can be in the range of zero to 100 and are expressed as a percentage, and a higher percentage indicates more accuracy in the accreditation of bank customers. To calculate indicators such as accuracy, it is necessary to calculate true positive, true negative, false positive, and false negative, which is represented by T_p, T_n, F_p, and F_n, respectively.

3. Results & Discussions

After introducing the data set in experiments and implementations, and then the implementation parameters, the evaluation indicators, and analysis of the proposed method are mentioned.

3.1. Output sample of the proposed method

To validate customers using the proposed method, two output samples are shown in Fig. 5, and 6, respectively. Fig. 5 shows the average error of customer classification and validation versus the repetition, and Fig. 6 illustrates the accuracy of customer classification and validation versus repetition. It can be seen that the customer authentication error decreases versus the repetition, and the accuracy increases, which shows the role of the swordfish optimization algorithm. The analysis of the proposed algorithm demonstrates that the classification and validation error of bank customers declines versus more iteration of the swordfish optimization algorithm. Table 1 shows the comparison of the proposed method with a population size of 10 with data mining methods in customer validation. It is observed that the average error, accuracy, sensitivity, and precision indices for multilayer artificial neural network (MLP), support vector machine (SVM), decision tree (DT), and random forest (RF).

Table 1 shows that the support vector machine has less MAE for customer validation than the mentioned methods consist of multilayer artificial neural network, and random forest decision tree that its value is about 30% less other methods. This reduction is due to the optimal choice of weight and bias. Also, the accuracy of machine learning methods comprises random forest, decision tree, support vector machine, and multilayer artificial neural network are 80.50%, 80.05%, 80.93%, and 81.58%, respectively, and the multilayer artificial neural network has the highest

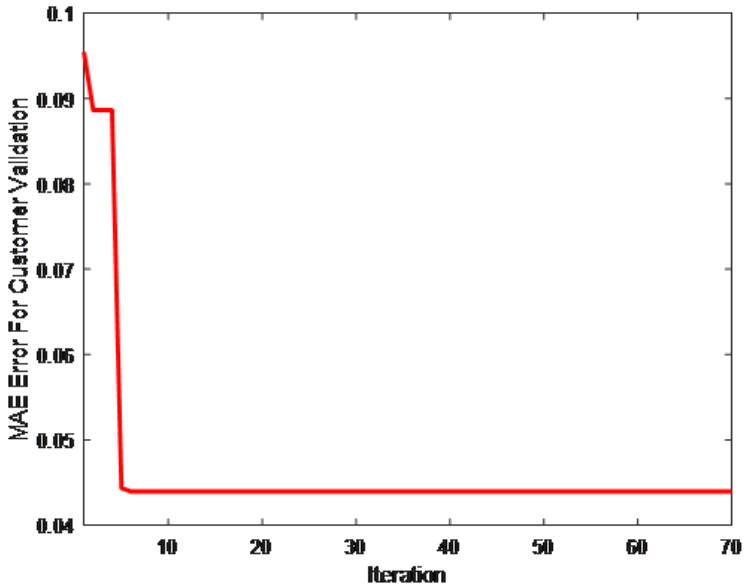

Figure 4. Validation error of bank customers in the proposed method versus the repetition with a population size of 10

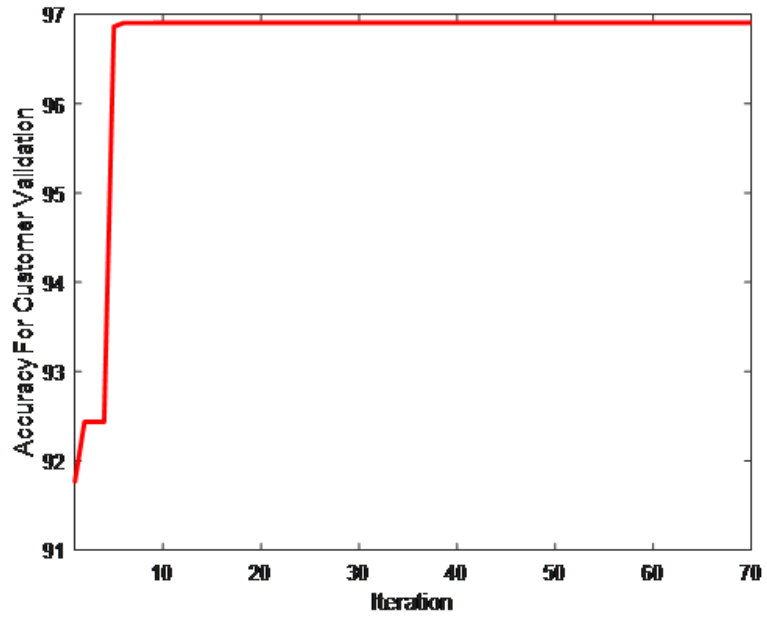

Figure 5. Validation accuracy of bank customers in the proposed method versus the repetition with a population size of 10

Table 1. The proposed methods for validating customers

Methods	MAE	Ac.(%)	Sen. (%)	Prec. (%)
MLP	0.274	81.58	81.6	80
SVM	0.191	80.93	80.9	79.2
DT	0.269	80.05	80.1	78
RF	0.270	80.5	80.5	78.2

accuracy for customer authentication. The sensitivity of random forest, decision tree, support vector machine, and multilayer artificial neural network are 80.50%, 80.10%, 80.90%, and 81.60%, respectively. Comparisons show that the multilayer artificial neural network has the highest sensitivity index for customer validation. Finally, the precision for the random forest, decision tree, support vector machine, and multilayer artificial neural network are 78.20%, 78%, 79.20%, and 80%, respectively. Therefore, the multilayer artificial neural network owns a higher precision index and has a greater ability to place the sample that was abnormal in the category of abnormal and show money laundering and non-credit of bank customers. According to the obtained results, the multilayer artificial neural network has higher accuracy, sensitivity, and precision index in the validation of bank customers than the machine learning methods such as random forest, decision tree, support vector machine. The multilayer artificial neural network is more successful as a basic method in the accuracy, sensitivity, and precision index than random forest, decision tree, support vector machine and performs customer accreditation more accurately.

4. Conclusions

In this paper, the machine learning algorithm was employed to classify model of bank customers' validation in order to reduce the validation recognition error. To this end, the machine learning algorithm was trained using a banking data set and then tested. The obtained results show that accuracy, sensitivity, and precision of the multilayer artificial neural network in customer validation in a population of 10 weight is 81.58%, 81.6%, and 80%, respectively, which has higher sensitivity, and accuracy than random forest methods, decision tree, support vector machine. Therefore, the multilayer artificial neural network has the best performance in terms of accuracy index for customer validation compared to other mentioned methods.

Conflict of interest

The authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest, or non-financial interest in the subject matter or materials discussed in this manuscript.

Acknowledgments

No applicable.

References

[1] Maitlo, G.M., Kazi, Z.H., Khaskheley, A. and Shaikh, F.M., "Factors that influence the adoption of online banking services in Hyderabad", International Journal of Economics & Management Sciences, 2015, 4(1), pp.1-10.

[2] Mousavian, S.J. and Ghasbeh, M.J.,"Investigation of Relationship between E-Banking Industry Risks and Electronic Customer Relationship Management (E-CRM)". MAYFEB Journal of Business and Management, 2017, 2.

[3] Sardana, S. and Bajpai, V.N., "E-banking service quality and customer satisfaction: an exploratory study on India". International Journal of Services and Operations Management, 2020, 35,(2), pp.223-247.

[4] Khedmatgozar, H.R. and Shahnazi, A., " The role of dimensions of perceived risk in adoption of corporate internet banking by customers in Iran", Electronic Commerce Research, 2018, 18(2), pp.389-412.

[5] Ojukwu-Ogba, N.E. and Osode, P.C., "A Critical Assessment of the Enforcement Regime for Combatting Money Laundering in Nigeria", African Journal of International and Comparative Law, 2020, 28(1), pp.85-105.

[6] González-Carrasco I, Jiménez-Márquez JL, López-Cuadrado JL, Ruiz-Mezcua B. "Automatic detection of relationships between banking operations using machine learning". Information Sciences, 2019, 485, pp.319-346.

[7] Sanz-Barbero B, Gómez AR, Ayala A, Recio P, Sarriá E, Díaz-Olalla M, Zunzunegui MV., "Impact of self-reported bank fraud on self-rated health, comorbidity and pain"

International journal of public health, 2020, 65(2), pp.165-174.

[8] Teichmann, F., "Recent trends in money laundering", Crime, Law and Social Change, 2020, 73(2), pp.237-247.

[9] Wang, Y., Wang, L. and Yang, J., " Egonet based anomaly detection in E-bank transaction networks. In IOP Conference Series", Materials Science and Engineering, IOP Publishing, 2020, Vol. 715, No. 1, p. 012038).

[10] Didimo, W., Grilli, L., Liotta, G., Menconi, L., Montecchiani, F. and Pagliuca, D., "Combining network visualization and data mining for tax risk assessment", IEEE Access, 2020, 8, pp.16073-16086.

[11] Sariannidis, N., Papadakis, S., Garefalakis, A., Lemonakis, C. and Kyriaki-Argyro, T., "Default avoidance on credit card portfolios using accounting, demographical and exploratory factors: decision making based on machine learning (ML) techniques", Annals of Operations Research, 2020, 294(1), pp.715-739.

[12] El-Banna, M.M., Khafagy, M.H. and El Kadi, H.M., "Smurf Detector: a Detection technique of criminal entities involved in Money Laundering", In 2020 International Conference on Innovative Trends in Communication and Computer Engineering (ITCE), 2020, (pp. 64-71). IEEE.

[13] Han, J., Huang, Y., Liu, S. and Towey, K., "Artificial intelligence for anti-money laundering: a review and extension", Digital Finance, 2020, 2(3), pp.211-239.

[14] Tadapaneni, N.R., "Artificial Intelligence in Finance and Investments".

[15] Chandradeva, L.S., Amarasinghe, T.M., De Silva, M., Aponso, A.C. and Krishnarajah, N.," Monetary Transaction Fraud Detection System Based on Machine Learning Strategies", In Fourth International Congress on Information and Communication Technology, Springer, Singapore, 2020, pp. 385-396.

[16] Fathi, M., Nemati, M., Mohammadi, S.M. and Abbasi-Kesbi, R.," A machine learning approach based on SVM for classification of liver diseases", Biomedical Engineering: Applications, Basis and Communications, 2020, 32(03), p.2050018.

[17] Khine, A.A. and Khin, H.W.," Credit Card Fraud Detection Using Online Boosting with Extremely Fast Decision Tree". In 2020 IEEE Conference on Computer Applications (ICCA), IEEE., 2020, pp. 1-4.

[18] Dong, M., Yao, L., Wang, X., Benatallah, B., Huang, C. and Ning, X., "Opinion fraud detection via neural autoencoder decision forest". Pattern Recognition Letters, 2020, 132, pp.21-29.

[19] Yan, C., Li, Y., Liu, W., Li, M., Chen, J. and Wang, L.,"An artificial bee colony-based kernel ridge regression for automobile insurance fraud identification". Neurocomputing, 2020, 393, pp.115-125.

[20] Abbasi-Kesbi, R., Memarzadeh-Tehran, H. and Deen, M.Jamal : "Technique to Estimate the Human Reaction Time Based on Visual Perception", Healthcare Technology Letters, 4, (2), 2017, pp. 73-77

[21] Abbasi-Kesbi, R., Asadi, Z. and Nikfarjam, A., "Developing a wireless sensor network based on a proposed algorithm for healthcare purposes", Biomedical engineering letters, 2020, 10(1), pp.163-170.

[22] Abbasi-Kesbi, R., Valipour, A. and Imani, K., "Cardiorespiratory system monitoring using a developed acoustic sensor", Healthcare technology letters, 2018, 5(1), pp.7-12.

[23] Abbasi-Kesbi, R. and Nikfarjam, A., "A miniature sensor system for precise hand position monitoring", IEEE Sensors Journal, 2018, 18(6), pp.2577-2584.

[24] Abbasi-Kesbi, R., Nikfarjam, A. and Memarzadeh-Tehran, H.: "A Patient-Centric Sensory System for In-Home Rehabilitation", IEEE Sensors Journal, 2017, 17, (2), p. 524-533.

[25] Onwubiko, C., "Fraud matrix: A morphological and analysis-based classification and taxonomy of fraud". Computers and Security, 2020, 96, p.101900.

[26] https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients\#

[27] Chai, T. and Draxler, R.R., "Root mean square error (RMSE) or mean absolute error (MAE)", Geoscientific Model Development Discussions, 2014, 7(1), pp.1525-1534.

[28] Wang, C., Yuan, H., Duan, Z. and Xiao, D., "Integrated multi-ISE arrays with improved sensitivity, accuracy and precision", Scientific reports, 2017, 7(1), pp.1-10.