Artificial Intelligence and Social Computing, Vol. 122, 2024, 52–61
https://doi.org/10.54941/ahfe1004641

Vector Result Rate (VRR): A Novel Method for Fraud Detection in Mobile Payment Systems

Arman Daliri[1], Mahdieh Zabihimayvan[2], and Kiarash Saleh[1]

[1]Department of Computer Engineering, Karaj Branch, Islamic Azad University, Karaj, Iran
[2]Department of Computer Science, Central Connecticut State University, New Britain, CT, USA

ABSTRACT

Mobile payment systems are becoming more popular due to the increasing number of smartphones, attracting fraudsters' attention. Therefore, existing researchers have developed various fraud detection methods using supervised machine learning. However, sufficient labeled data are rarely available, and their detection performance is negatively affected by severe class imbalance in financial fraud data. This study aims to propose a new model entitled Vector Result Rate (VRR) for fraud detection based on deep learning while considering the economic consequences of fraud detection systems. The proposed framework is experimentally implemented on a large dataset containing more than six million mobile phone transactions. A comparative evaluation of existing machine learning methods designed to model unbalanced data and detect outliers is performed for the comparison. The results show that the VRR achieves the best results by integrating several classification algorithms with supervision and classifiers regarding standard classification criteria.

Keywords: Fraud detection, Machine learning, Deep learning, Unbalanced data

INTRODUCTION

Fraud detection in mobile payment applications is one of the crucial issues in the field of electronic payment security (Burri et al., 2023). Due to the growing trend of electronic payments and the widespread use of mobile phones, fraud in this area is also increasing. To detect fraud in these applications, various algorithms try to identify fraudulent patterns by analyzing the behavior of users and their transactions (Burri et al., 2023). These algorithms can use different methods, such as neural networks, decision trees, and machine learning. The importance of fraud detection in these applications is vital in two ways. First, electronic payment fraud can cause financial loss to payment companies and customers. Then, these frauds can reduce users' trust in the electronic payment system and thus reduce the use of these systems (Koskelainen et al., 2023). Fraud detection in mobile payment applications generally increases electronic payment security and users' trust in these systems.

Supervised learning is one of the methods used in fraud detection in mobile bank software (Ivanyuk, 2023). Machine learning algorithms identify fraudulent patterns in supervised learning without requiring labeled data or human supervision. Supervised machine learning algorithms learn the typical patterns of user behavior by using input data such as the behavior of users in mobile bank software (Ivanyuk, 2023). Using these patterns, transactions that deviate from standard patterns are known and identified as fraudulent. Supervised learning algorithms can be implemented using different methods, including neural networks and classification algorithms (Alimoradi et al., 2022). These methods can identify fraudulent patterns using techniques such as principal component analysis. The use of supervised learning methods in fraud detection in mobile bank software has the advantage of high accuracy due to the need for labeled data and human supervision (Alimoradi et al., 2022). Also, this method can significantly improve the accuracy of fraud detection due to the ability to adapt to changes in user behavior (Daliri et al., 2024).

Research in the field of fraud detection in banking applications can face challenges, some examples of which are given in this section. One of the challenges is the need for educational data. To train fraud detection systems, the need big data that includes different fraud patterns (Strelcenia and Prakoonwit, 2023). However, in some cases, insufficient training data is available for the fraud detection system, which can reduce the system's accuracy (Daliri et al., 2022). Fraudsters are constantly trying to update their algorithms to bypass fraud detection systems (Paladini et al., 2023a). For this reason, fraud detection systems must be updated to adapt to changes in fraud algorithms (Paladini et al., 2023a). The third challenge is security problems. Fraud detection systems must be robust in terms of security to prevent fraudsters' attacks (Paladini et al., 2023b). However, in some cases, fraud detection systems may be attacked and fail due to weak security (Paladini et al., 2023b). The last challenge under consideration is legal restrictions. In some countries, fraud detection systems in banking applications may face legal restrictions (Hendieh et al., 2023). For this reason, companies operating in this field must comply with data privacy and security laws (Hendieh et al., 2023).

In this article, to help this field of research and to help different communities, an attempt has been made to provide a detailed framework for fraud detection in mobile bank systems. Considering that the previous methods have performed well, this research presents a new framework for fraud detection according to the expert points. Generally, this framework deals with high accuracy for fraud classification and detection by pre-processing data and using innovation in pre-processing. The proposed framework starts by balancing the database and choosing the best balancer algorithm for the type of data that is read and then examines the most essential features of the database. Finally, after processing the data, it implements deep algorithms and goes through the framework of training and testing. Among the most important contributions of this research, the following can be mentioned:

- Presenting a new hybrid framework that combines three artificial intelligence methods and is an artificial intelligence for artificial intelligence.

- Introducing a new method for choosing the balancing algorithm of the target feature of the database.
- Providing a long-term and implementable solution to other forecasting and diagnosis issues in the intelligent banking industry.
- Using a PaySim benchmark dataset with more than 6 million mobile payment transactions, it is demonstrated that the proposed fraud detection framework not only outperforms state-of-the-art fraud detection methods in terms of detection accuracy.

In this research, firstly, in section "Related works", related works have been presented. Then, in section "Vector Result Rate (VRR) For Fraud Detection", the proposed method for fraud detection in mobile bank systems has been introduced. The results are presented and compared with the previous techniques to check the performance of the proposed method in section "Experimental results". Finally, in section "Conclusion", conclusions and future works are discussed. 2003).

RELATED WORKS

Fraud detection in mobile bank systems with the help of machine learning algorithms is one of the important applications of machine learning in banking (Onyema et al., 2023). Algorithms detect fraud in new transactions using data collected from previous transactions (Onyema et al., 2023). The history of fraud detection in banking systems with the help of machine learning algorithms dates back to the 1990s (Ricketts, 2023). At that time, simple algorithms were used to detect fraud in bank transactions. However, with the advancement of technology and the increase in the volume of bank transactions, the need for more advanced algorithms to detect fraud was felt (Djomadji et al., 2023). In recent years, with the advancement of machine learning methods, more complex algorithms have been developed to detect fraud in bank transactions. These algorithms can detect deception in bank transactions with higher accuracy by using classification methods in machine learning (Onyema et al., 2023).

The article (da Silva et al., 2023) shows that to detect fraud in financial systems, it is crucial to review and analyze the data carefully. In this study, the PaySim dataset has been used, and image data analysis has shown how image data analysis can reveal the weak points and strengths of the data. Also, the importance of identifying abnormal issues and dangerous potentials in the data has been investigated. The results show that image data analysis is one of the basic steps in identifying fraudulent activities in monetary transactions (da Silva et al., 2023).

According to the 2022 Cybercrime Report, the number of complaints and financial damages from 2018 to 2022 show $27.6 billion in 3.26 million complaints. Technology is being developed by institutions interested in reducing cybercrime, and researchers are collaborating with them to implement fraud prediction systems (Bezerra Junior, 2023). Machine learning and deep learning are applied in various studies to understand and learn how to prevent fraudulent transactions in real financial networks by using past transactions.

In this paper (Lopez-Rojas et al., 2016), it is proposed to use a modified version of the cross-validation technique of the total set approach on a synthetic dataset of wire legs and send it to a neural network model. Its results with a five-fold division method of 20% of the data set of each sheet is compared with the same model and then with random forest, logistic regression, and Ada-bust machine learning algorithms. The measures applied to evaluate the performance of the models are precision, accuracy, recall, and F1 score (Bezerra Junior, 2023).

VECTOR RESULT RATE (VRR) FOR FRAUD DETECTION

In this research, we have tried to detect fraud by presenting a combination of statistical and machine-learning methods. The framework has been explained step by step to understand the proposed method better. For this purpose, an overview of the proposed framework is given in Figure 1, and then each step is explained according to the figure. In Figure 1, six general steps for fraud detection are shown, which are examined to reach the diagnosis. In general, the proposed method is described in two subsections. First, in section DATABASE PREPROCESSING, information about data processing is given. In the last part, in sub-section FRAUD CLASSIFICATION, classification methods and an explanation of the process with flowcharts have been discussed.

DATABASE PREPROCESSING

First step to third one is implemented in this part of the proposed framework. The proposed framework has been implemented using the PaySim database provided in the article (Lopez-Rojas et al., 2016). PaySimhas simulated 743 time series representing thirty days of real-time data. One thousand fraudsters with a three percent probability of committing fraud were entered at each stage to introduce fraudulent behavior into the system. A total of 6,362,620 mobile phone transactions were made in this data set, of which 8,213 were fraudulent. Table 1 presents the descriptive statistics of the dataset.

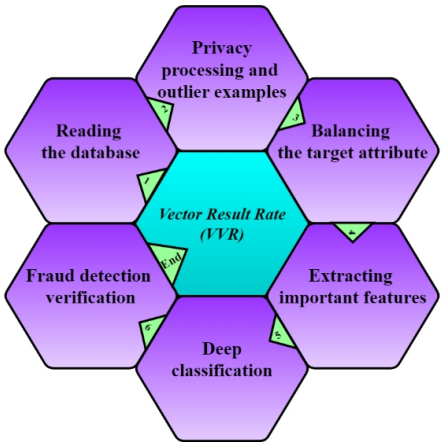

Figure 1: Summary of the proposed fraud detection framework.

Table 1. Descriptive statistics of the PaySim data set.

Attribute	Average Value/Range
Step	1–743
Type of transaction	cash (35%) cash (34%) Transfer and debt (31%)
Amount of transaction	180 thousand
Customer name	million unique values 6.35
Initial balance	834 thousand
New Balance	855 thousand
Recipient name	2.72 million unique values
The initial balance of the recipient	1.1 million
The new balance of the recipient	1.22 million
Fraud	(million legitimate 6.36)0 (8.2K cheat) 1

As this framework is intended to work with different datasets in the future, any personal information such as identification numbers, national addresses, telephone numbers, or the like is removed from the datasets in this section. In addition, data figures are retrieved for visual analysis of features. Duplicate samples are extracted to prepare the data for training. Parts that have outliers are also identified after extraction. Finally, the two subjects are merged, and duplicates are removed. The actual data set has a high probability of containing incorrect data. For this reason, all incorrect data has been removed from this database. After that, the data balance is applied after removing the existing null data. If the data set and its target are unbalanced, data balancing techniques are implemented before moving to the next step.

After preparing the database, we use balancing methods to balance the data. Several balancing methods have been proposed and used in various machine-learning studies. In order to make a correct decision in choosing a suitable balancing method, a new way is presented, the Vector Result Rate (VRR). Several balanced algorithms have been implemented to implement the new proposed plan. Finally, the correct performance of the vector result rate (VRR) method has been shown by extracting the results and comparing them with the desired number in a new way. The counter balancers implemented for validation are Random Affine Shadow sampling (LORAS) (Bej et al., 2021), synthetic minority over-sampling technique (SMOTE) (Pears et al., 2014), Random Oversampling (de Morais and Vasconcelos, 2019), and synthetic minority over-sampling technique with Extended nearest neighbor (SMOTE-ENN) (Cao and Wang, 2011).

The rate vector works based on two numerical criteria to select the best balancing method from a set of balancing algorithms. For each candidate, the values of these two criteria are calculated and used to construct a vector. Then, the length of each vector is calculated and compared with the length of other algorithms. The two criteria used in the vector result rate are set so that the best balancing algorithm is the algorithm with the smallest vector length. The two criteria used in the vector result rate are the Calinski-Harabasz score

(Wang and Xu, 2019) and the silhouette (Shahapure and Nicholas, 2020), which are as follows:

Calinski-Harabasz uses the Keymeans algorithm to predict suitable classes in a data set (Wang and Xu, 2019). The Calinski-Harabasz index is the ratio of cluster dispersion to the sum of intracluster dispersion for each cluster. When the clusters are dense, the value of this criterion is higher. Equation 1 shows the mathematical formula of this metric. Assume that E has size n_E and is clustered into k clusters. Calinski-Harabasz, denoted as s, is the ratio between-cluster to within-cluster dispersion. $tr(B\_k)$ traces the scatter matrix between the defined cluster and relation 2. $tr(W\_k)$ is defined as the within-cluster scatter matrix that traces the relation 3. C_i is the center of cluster i, and n_q is the number of points in that cluster q.

$$s = \frac{tr(B_k)}{tr(W_k)} \times \frac{n_E - k}{k - 1} \tag{1}$$

$$W_k = \sum_{q=1}^{k} \sum_{x \in C_q} (x - C_q)(x - C_q)^T \tag{2}$$

$$B_k = \sum_{q=1}^{k} n_q \left(C_q - C_E\right) \left(C_q - C_E\right)^T \tag{3}$$

The silhouette score uses the K-means algorithm to define the appropriate classes of a data set (Shahapure and Nicholas, 2020). The value of this metric ranges from -1 to $+1$, and a maximum value means that all classes are distinguishable from each other. The mathematical formula of the silhouette score is shown in Equation 4, where s is the silhouette score, a is the average distance between each sample and other points in the same class, and b is the average distance between a sample and all other points in the next class that is the closest cluster.

$$s = \frac{b - a}{\max(a.b)} \tag{4}$$

FRAUD CLASSIFICATION

After the data exits the VRR, all the classification algorithms are repeated once. The evaluation and scoring system are implemented. The algorithms implemented in this section are Convolutional neural network (CNN) (Shah et al., 2023), Recurrent neural network (RNN) (Khanduzi and Sangaiah, 2023), and Multilayer perceptron (MLP) (Kumar and Yadav, 2011). All initial parameters for each classifier are set to default values. The criteria that are used to evaluate the performance of the classifiers. The results are compared with accuracy, precision, recall, and F1 score criteria. Equations 5, 6, 7, 8 P indicates the size of the positive class, N indicates the size of the hostile class, TP is the number of samples in the positive class, and TN is the number of samples in the negative class. Class, FP is the number of samples that are falsely positive, and FN is the number of samples that are falsely negative in

the class.

$$Accuracy = \frac{TP + TN}{P + N} \tag{5}$$

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

$$F1 - score = \frac{(2\,(TP))}{(2\,(TP)) + Recall} \tag{8}$$

The primary purpose of this method is to select balancing algorithms automatically. By implementing this algorithm, researchers will not need extensive research. They will only spend a small amount of time choosing the algorithm because this algorithm aims to solve this time-consuming problem. Finally, with the design of this algorithm, fraud is predicted very quickly and correctly, and it is a great help in the banking field. The complete steps of this algorithm are as follows:

1. In this algorithm, the data is first read, and the information is processed according to the description of subsection DATABASE PREPROCESSING.
2. Balancing algorithms are implemented, and one is selected according to the vector rate.
3. The best algorithm and score will be displayed if satisfied.

Experimental Results

In this section, the performance of the new method is examined. Analyzes are provided to prove the correct performance of the new criteria. Then, the methods with raw data and algorithms are compared separately. This section aims to improve the results by using the new algorithm. First, the vector rate metric is presented in subsection THE VECTOR RATE METRIC. Then, in next subsection, THE RESULTS OF CLASSIFICATION ALGORITHMS are presented. Finally, the comparison of the performance of the source algorithms with the proposed framework is given.

THE VECTOR RATE METRIC

Table 2 lists the vector lengths for each balancing method after applying VRR. Based on the results, LORAS has the lowest length area, equal to 0.6297, and SMOTE ranks second by a slight difference. ROS has the most significant area with an area of 0.7001, and SMOTE-ENN has lengths close to the maximum. We use LORAS as a balancer for further experiments because it has the minimum vector rate length.

To further analyse the performance of the VRR metric and show the relationship between the values of the vector rate regions and the accuracy of the classifiers, Table 3 shows how an increase in the vector rate level can display an increase in the classification accuracy. In other words, the table shows how the average classification accuracy changes as the area of the triangle increases by 0.1. As the values show, any increase in the vector decreases the classification accuracy.

Table 2. Vector length for the balancers under review.

Balancing algorithm	LORAS	SMOTE	ROS	SMOTE-ENN
Vector Result Rate	0.6297	0.6754	0.7001	0.6938

Table 3. Threshold of vector change in case of decrease of vector result rate criterion.

Algorithm / Tolerance	0.6	0.7
Convolutional neural network (CNN)	98%	85%
Recurrent neural network (RNN)	95%	83%
Multilayer perceptron (MLP)	92%	81%

THE RESULTS OF CLASSIFICATION ALGORITHMS

The raw data used in the experiments includes 6362620 transactions, of which 1000 are in the fraudulent group, and the rest are in the healthy transactions group. Table 3 shows the evaluation criteria, F1 score, precision, and recall for four different equalizers and three classification methods. The first row shows the performance of the classifiers on the original data, which is unbalanced. The results show better performance for the majority class, which are people and regular transactions. However, this work focuses on providing a precise classification for each fraudulent and healthy case.

The rest of the table shows the classification performance when the four balancing methods balance the data. All moderators affect the evaluation results, especially the minority class, the cheating group. The results show that the classification of balanced data by LORAS leads to better performance than others. CNN has the highest F1 score compared to other classification methods. For other evaluation criteria, the results of these two classifiers are almost close to each other. However, the overall performance of the CNN method is higher than the others for both fraudulent and healthy classes.

Table 4. F1 score, precision (Pr), and recall (R) for three classifiers on unbalanced and balanced data.

Balancer	Classifier	F1 of the right group	F1 of the fraud group	Pr of the right group	Pr of the fraud group	R of the right group	R of the fraud group
Raw data	CNN	85%	15%	75%	62%	98%	9%
	RNN	88%	53%	82%	75%	95%	42%
	MLP	85%	58%	84%	59%	86%	56%
LORAS	CNN	**98%**	**97%**	**98%**	**97%**	**98%**	**97%**
	RNN	**95%**	**94%**	**95%**	**94%**	**95%**	**94%**
	MLP	**91%**	**89%**	**91%**	**89%**	**91%**	**89%**
SMOTE	CNN	92%	89%	90%	89%	90%	89%
	RNN	91%	89%	91%	89%	91%	89%
	MLP	89%	88%	89%	88%	89%	88%
ROS	CNN	91%	89%	91%	89%	91%	89%
	RNN	77%	76%	76%	77%	78%	75%
	MLP	74%	67%	67%	76%	82%	59%
S-ENN	CNN	89%	88%	89%	88%	89%	88%
	RNN	80%	82%	80%	81%	79%	82%
	MLP	75%	73%	69%	80%	81%	68%

CONCLUSION

This research presents a new fraud detection framework to predict healthy and fraudulent transactions. First, information about fraud detection methods is given, and then the used methods are explained. The proposed method provides a new criterion suitable for choosing the balancing algorithm. By combining this criterion in the learning process, the possibility of accurate prediction and diagnosis has been realized. Considering the importance of fraud detection under study and the success of the presented algorithm, future works can be theorized by this framework.

According to the work done, several future works can be implemented for this article. Big data can be used to evaluate this task and to implement additional algorithms, either classifiers or balancers. From the ideas of the future implementation of this algorithm, it can be thought of providing a web version of the framework for predicting fraud based on user input.

REFERENCES

Alimoradi, M., Zabihimayvan, M., Daliri, A., Sledzik, R., Sadeghi, R., 2022. Deep Neural Classification of Darknet Traffic, in: Cortés, A., Grimaldo, F., Flaminio, T. (Eds.), Frontiers in Artificial Intelligence and Applications. IOS Press. https://doi.org/10.3233/FAIA220323

Bej, S., Davtyan, N., Wolfien, M., Nassar, M., Wolkenhauer, O., 2021. LoRAS: An oversampling approach for imbalanced datasets. Mach. Learn. 110, 279–301. https://doi.org/10.1007/s10994-020-05913-4

Burri, S. R., Kumar, A., Baliyan, A., Kumar, T. A., 2023. Transforming Payment Processes: A Discussion of AI-Enabled Routing Optimization, in: 2023 2nd International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN). IEEE, pp. 1–7.

Camilo da Silva, M., Tavares, G. M., Gritti, M. C., Ceravolo, P., Barbon Junior, S., 2023. Using Process Mining to Reduce Fraud in Digital Onboarding. FINTECH 2, 120–137.

Cao, Q., Wang, S., 2011. Applying over-sampling technique based on data density and cost-sensitive SVM to imbalanced learning, in: 2011 International Conference on Information Management, Innovation Management and Industrial Engineering. IEEE, pp. 543–548.

Daliri, A., Alimoradi, M., Zabihimayvan, M., Sadeghi, R., 2024. World Hyper-Heuristic: A novel reinforcement learning approach for dynamic exploration and exploitation. Expert Syst. Appl. 244, 122931.

Daliri, A., Asghari, A., Azgomi, H., Alimoradi, M., 2022. The water optimization algorithm: A novel metaheuristic for solving optimization problems. Appl. Intell. 52, 17990–18029. https://doi.org/10.1007/s10489-022-03397-4

de Morais, R. F., Vasconcelos, G. C., 2019. Boosting the performance of over-sampling algorithms through under-sampling the minority class. Neurocomputing 343, 3–18.

Djomadji, E. M. D., Basile, K. I., Christian, T. T., Djoko, F. V. K., Sone, M. E., 2023. Machine Learning-Based Approach for Identification of SIM Box Bypass Fraud in a Telecom Network Based on CDR Analysis: Case of a Fixed and Mobile Operator in Cameroon. J. Comput. Commun. 11, 142–157.

Hendieh, J., Schneider, M., Sakr, T., 2023. Fraud Detection and Prevention. Middle-East J. Sci. Res. 31, 44–52.

Ivanyuk, V., 2023. Forecasting of digital financial crimes in Russia based on machine learning methods. J. Comput. Virol. Hacking Tech. https://doi.org/10.1007/s11416-023-00480-3

Khanduzi, R., Sangaiah, A. K., 2023. An efficient recurrent neural network for defensive Stackelberg game. J. Comput. Sci. 67, 101970.

Koskelainen, T., Kalmi, P., Scornavacca, E., Vartiainen, T., 2023. Financial literacy in the digital age—A research agenda. J. Consum. Aff. 57, 507–528. https://doi.org/10.1111/joca.12510

Kumar, M., Yadav, N., 2011. Multilayer perceptrons and radial basis function neural network methods for the solution of differential equations: A survey. Comput. Math. Appl. 62, 3796–3811.

Lopez-Rojas, E., Elmir, A., Axelsson, S., 2016. PaySim: A financial mobile money simulator for fraud detection, in: 28th European Modeling and Simulation Symposium, EMSS, Larnaca. Dime University of Genoa, pp. 249–255.

Onyema, J. C., Betrand, C. U., Benson-Emenike, M., 2023. Machine Learning Credit Card Fraud Detection System. Appl. Sci. Res. Period. 1, 19–28.

Paladini, T., Bernasconi De Luca, M., Carminati, M., Polino, M., Trovò, F., Zanero, S., 2023a. Advancing Fraud Detection Systems Through Online Learning, in: De Francisci Morales, G., Perlich, C., Ruchansky, N., Kourtellis, N., Baralis, E., Bonchi, F. (Eds.), Machine Learning and Knowledge Discovery in Databases: Applied Data Science and Demo Track, Lecture Notes in Computer Science. Springer Nature Switzerland, Cham, pp. 275–292. https://doi.org/10.1007/978-3-031-43427-3_17

Paladini, T., Monti, F., Polino, M., Carminati, M., Zanero, S., 2023b. Fraud Detection Under Siege: Practical Poisoning Attacks and Defense Strategies. ACM Trans. Priv. Secur. 3613244. https://doi.org/10.1145/3613244

Pears, R., Finlay, J., Connor, A. M., 2014. Synthetic Minority Over-sampling TEchnique (SMOTE) for Predicting Software Build Outcomes.

Ricketts, T., 2023. Speech Recognition Application With Tone Analyzer (PhD Thesis). Alabama Agricultural and Mechanical University.

Rocha Bezerra Junior, J. B., 2023. Overcoming Imbalanced Class Distribution and Overfitting in Financial Fraud Detection: An Investigation Using A Modified Form of K-Fold Cross Validation Approach to Reach Representativeness (PhD Thesis).

Shah, A., Shah, M., Pandya, A., Sushra, Rajat, Sushra, Ratnam, Mehta, M., Patel, Keyur, Patel, Kaushal, 2023. A Comprehensive Study on Skin Cancer Detection using Artificial Neural Network (ANN) and Convolutional Neural Network (CNN). Clin. EHealth.

Shahapure, K. R., Nicholas, C., 2020. Cluster quality analysis using silhouette score, in: 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA). IEEE, pp. 747–748.

Strelcenia, E., Prakoonwit, S., 2023. A Survey on GAN Techniques for Data Augmentation to Address the Imbalanced Data Issues in Credit Card Fraud Detection. Mach. Learn. Knowl. Extr. 5, 304–329.

Wang, X., Xu, Y., 2019. An improved index for clustering validation based on Silhouette index and Calinski-Harabasz index, in: IOP Conference Series: Materials Science and Engineering. IOP Publishing, p. 052024.