Bibliometric Analysis and Visualization of Machine Learning-Based Credit Card Fraud Detection

1st Suhartono
Department of information engineering
UIN Maulana Malik Ibrahim Malang
Malang, Indonesia
suhartono@ti.uin-malang.ac.id

2nd Syahiduz Zaman
Department of information engineering
UIN Maulana Malik Ibrahim Malang
Malang, Indonesia
syahid@ti.uin-malang.ac.id

3rd Totok Chamidy
Department of information engineering
UIN Maulana Malik Ibrahim Malang
Malang, Indonesia
to2k2013@ti.uin-malang.ac.id

2024 International Conference on Information Technology Research and Innovation (ICITRI) | 979-8-3503-7621-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICITRI62858.2024.10699070

***Abstract*— Machine learning is often used in credit card fraud detection. Its ability to analyze large amounts of transaction data and identify patterns of fraudulent activity. The aim of this study is to provide insights into the research status, mapping process and annual topics of machine learning research on credit card fraud detection, and to provide relevant references for future research. Research phases with three approaches. The first approach is descriptive statistical analysis for data collection using Scopus web and Herzing's Publish or Perish. The second approach uses quantitative methods for citation analysis with a bibliometric approach using Vos Viewer application. The research findings are related to the analysis of credit card fraud research divided into three clusters, the first cluster is the research stream using machine learning in data mining, the second cluster is the research stream using machine learning to detect credit card fraud, the third cluster is the research stream using machine learning to classify credit card fraud. The second cluster can resolve the complexity of credit card transaction data and increase the accuracy of the fraud detection system, for the third cluster, we can build a more effective classification in resolving imbalanced data and limited transaction records. A growing research trend is in the third cluster, research related to the performance of credit card fraud classification based on unbalanced data and limited fraud data. Further research is recommended to examine the evolution of the literature on the use of machine learning to detect credit card fraud and to conduct comparative studies between the Scopus, Web of Science and ScienceDirect databases to expand the literature.**

Keywords*— Machine learning; detection; fraud; credit cards; analysis; bibliometrics; visualization

I. INTRODUCTION

Currently, research into credit card fraud detection is important because credit card usage is increasing every year, which is accompanied by an increase in fraudulent activity in credit card transactions. Implementing an efficient fraud detection system is mandatory for credit card issuing financial institutions to minimize credit card fraud, according to Raj & Portia [1]. Financial transactions are very important for economic growth. Therefore, prevention through credit card fraud detection serves to maintain public trust in credit card transactions [2]. Therefore, it is necessary to develop a fraud detection model for credit card transactions to minimize losses for credit card customers and maintain the integrity of credit card transactions [3].

Machine learning is a branch of artificial intelligence that focuses on developing techniques that enable systems to learn from data and make predictions or decisions without external programming [4]. In the context of developing credit card fraud detection models, machine learning is often used to analyse transaction patterns and detect suspicious activities [5]. The use of machine learning to detect credit card fraud has made significant progress in recent years. Various studies have proposed various machine learning methods such as deep learning, supervised learning algorithms, and ensemble methods to improve the detection of fraudulent credit card transactions [1-2] [5] [6-12]. Researchers have used various machine learning methods, including support vector machines, logistic regression, random forests, and gradient boosting. The aim of using different methods is to increase the effectiveness in identifying fraudulent activities [1-2] [11].

In addition, the use of deep learning, such as convolutional neural networks and deep neural networks, has been explored to improve the accuracy of fraud detection [5] [12]. The research community continues to explore innovations to further improve the accuracy of credit card fraud detection models through machine learning [13].

In this article, we explore the use of machine learning to detect credit card fraud. The aim is to provide a bibliometric analysis and visualization on this research topic. In bibliometric analysis we use tools such as the Vos Viewer application. This tool can provide an overview of the research findings in machine learning research on credit card fraud detection. Bibliometric analysis includes volume, connections, citations, productivity, quality, and emerging trends. This article also explains how to use related keyword combinations to search for relevant documents in the Scopus database.

This research also uses visualizations such as images and maps to present bibliometric analysis results including most productive countries, top authors, co-author productivity and network cluster distribution in the field of using machine learning to detect credit card fraud. The aim of this research is also to find out how much research there is on a particular topic, to provide an overview of the state of research, the mapping process and the annual themes, and to offer potential references for future research.

To answer the research objectives, we propose five research questions (RQ). RQ1: What is the trend in the number of publications investigating machine learning for credit card fraud detection? RQ2: Which scientific articles are most frequently cited? RQ3: What is the card? Development of scientific publications based on keywords? RQ4: Which institutions, countries, journals and authors are leaders in this area? RQ5: Which articles and topics are the most influential and trending?.

II. METHODOLOGY

This article uses open-source software for bibliometric analysis and visualization, namely Herzig's Publish or Perish, Vos Viewer and Scopus web application. The aim of bibliometric analysis is to map and analyze scientific publications. The advantage of bibliometric analysis is that it provides information that is concise, easy to use and tends to be objective.

The Vos Viewer applications are used to analyze citations, create citation networks between different studies on a research topic, analyze the performance of researchers, map fields to research topics, and perform network analysis to identify intellectual structures, conceptual frameworks, and relationships within research fields to understand topic. This research also determines the criteria for excluded articles.

Article selection was carried out using eligibility criteria. Eligibility criteria consist of inclusion and exclusion criteria. The criteria are: articles not in English will be excluded, articles not related to the research topic will be excluded, articles outside the agreed publication year will be excluded.

In addition, this research also determines the inclusion criteria and relevant articles. The inclusion criteria are articles that are written in English, articles that fit the agreed research year and articles that are related to the research topic. The complete literature source selection process is shown in Figure 1.

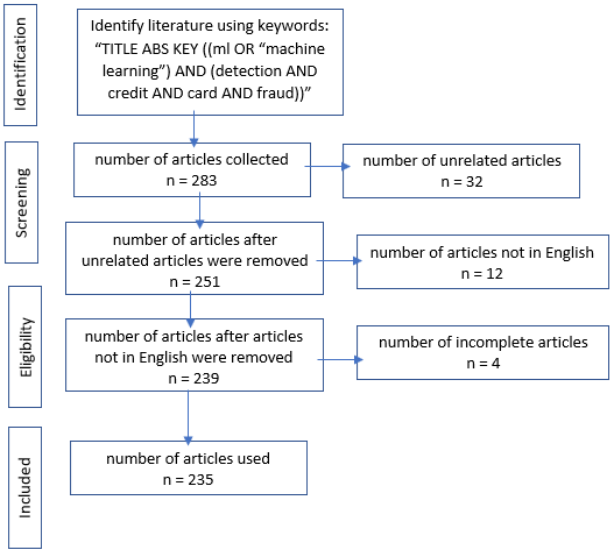

Fig. 1. Scientific Article Selection Process

This research uses a database obtained from Scopus web application. The database used was published between 2004 and 2024. The database includes different types of documents such as Articles (28.5%), Proceedings (60.0%), Books (0.9%), Books Section (3.8%), Review of Procedure (4.3%) and corrections (2.6%) as in Figure 2.

The collected database was used to answer five research questions related to the most prolific authors in the research topic, the annual growth of scientific publications, the collaboration between countries contributing to the research topic, the most published journals, the most cited scientists and the Concept relate to structural field. Keyword selection in bibliometric analysis research is to use the query "TITLE ABS KEY (ml OR "machine learning") AND (detection AND credit AND card AND fraud))". The keywords in this query are used to search for articles. Articles based on keywords relevant to the research topic. The results of running this query returned 283 articles.

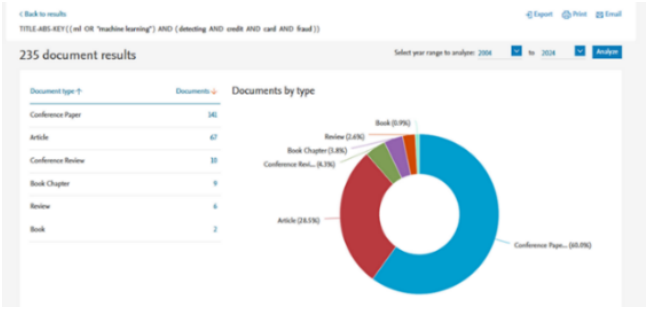

Fig. 2. Screenshot of Scopus web application based on document type

The following process filters articles based on relevance to the research topic area. Articles with unrelated scientific fields will be excluded. The number of excluded articles is 32. Articles that are not in English are excluded; the number of excluded articles is 12. Articles with incomplete and unclear information are excluded, the number of excluded articles is 4. Articles that relate to the selected research area The number of relevant articles is 235. Meanwhile, the qualitative evaluation process (descriptive analysis) was carried out using Herzing's Publish or Perish software and the analysis was carried out using Scopus web application. The quantitative evaluation process (citation analysis) is now carried out using the Vos Viewer application.

III. RESULTS

This research produced to a basic statistical analysis using Herzing's Publish or Perish software. The number of articles found was 235 articles, the publication date of the articles was between 2004 and 2024, and the number of citations was 3492 citations, as shown in Figure 3.

Fig. 3. Screenshot of Herzing's Publish or Perish application for basic statistical analysis based on article publication time from 2004 to 2024.

A. Annual article production

Figure 4 shows the trend in the number of publications per year, namely between 2004 and 2024. The results of trend analysis using the Scopus web application are as in Figure 1: The trend in the number of publications per year gradually decreases over time to, the publication period begins at the beginning of 2004 and reaches its peak in the years 2022 to

2023. The highest number of publications will be in 2023 with 66 publications. The lowest number of publications occurred in 2006 and 2007 with 2 publications each.

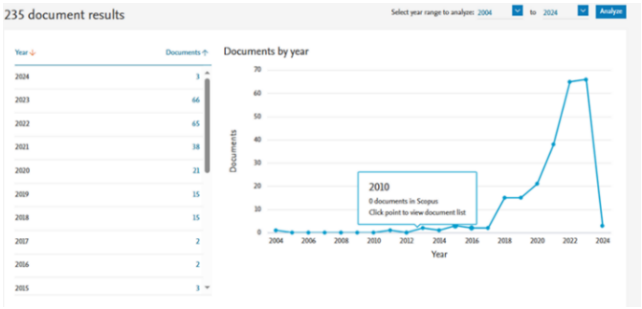

Fig. 4. Screenshot of the Scopus web application related to the evolution of the number of publications per year

B. Influential Countries, Institutions and Departments

The results of trend analysis using the Scopus web application are as shown in Figure 5. The trend analysis is done in the form of a histogram with respect to the ten countries that produce the most articles on the research topic, India is the first country with 103 articles, the second is the United States with 23 articles and the second is the United States with 23 articles. The third country is China with 19 articles.

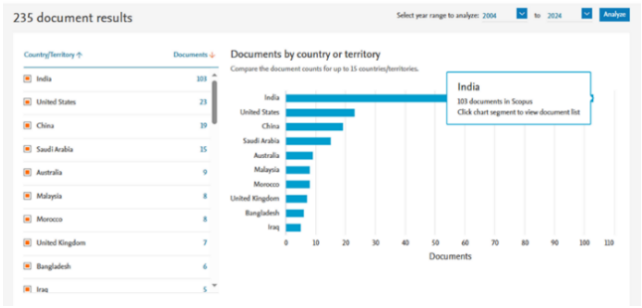

Fig. 5. Screenshot of the Scopus web application showing the top ten countries producing the most articles

Figure 6 shows a histogram of the ten institutions that produce the most articles on research topics. The National Key Research and Development Program of China ranks first with 4 articles, the second place is the National Natural Science Foundation of China with 4 articles, and the third place is King Abdulaziz University with 3 articles.

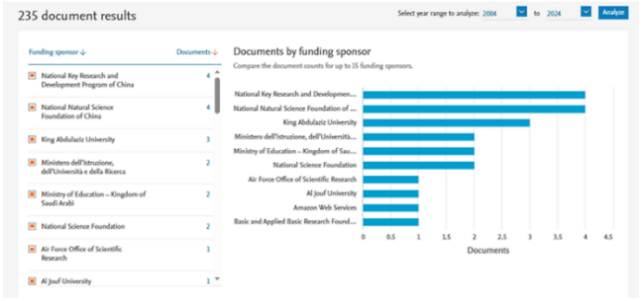

Fig. 6. Screenshot of the Scopus web application showing the ten institutions producing the most articles

Figure 7 shows a histogram of the ten departments that produce the most articles on research topics. First is Lovely Professional University with 5 articles, second is SRM Institute of Science and Technology with 4 articles, and third is Florida Atlantic University with 4 articles.

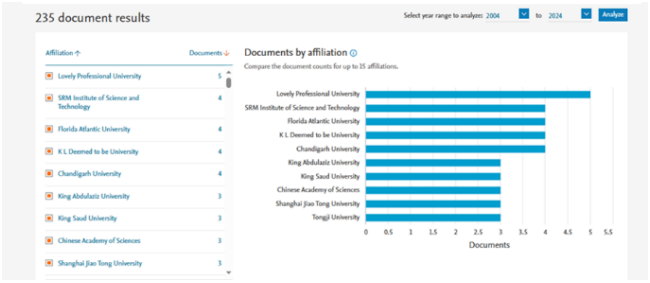

Fig. 7. Screenshot of the Scopus web application showing the ten departments producing the most articles

C. Most Influential Journals

Figure 8 shows a histogram of the three journals that produce the most articles on research topics. Lecture notes in networks and systems are the first with 9 articles, the second is the journal ACM International Conference Proceedings Series with 7 articles, and third is the proceedings of the AIP journal with 7 articles.

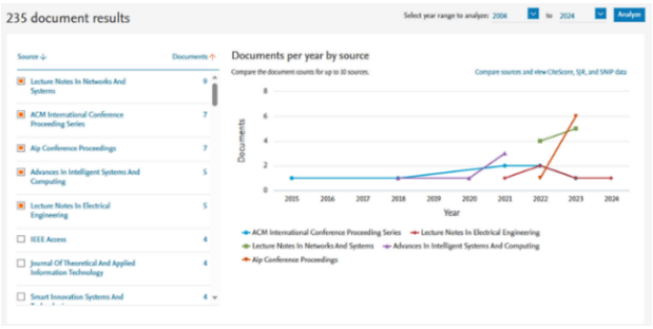

Fig. 8. Screenshot of the Scopus web application on the most influential journals in research.

D. The Most Influential Authors

Figure 9 shows a histogram of the author who wrote the most articles on the research topic. First and foremost is the author Khoshgoftaar T.M. with 4 articles, second is author Prusti D with 3 articles and third is author Khoshgoftaar T.M author Rath R.K. with 3 items.

Fig. 9. Screenshot of the Scopus web application on the most influential journals in research.

E. Co-writing network

The results of the collaborative authoring network analysis using the Vos Viewer application are shown in Figure 10. Of the 235 articles, there are 761 authors, collaborative articles are identified with a minimum of 2 authors and a maximum of 8 authors per document, collaborative publications based on 6 documents can be described in the form of a co-author network, two clusters of 6 collaborative documents were identified collaborative authors found, the first cluster in red with three authors, consisting of Khoshgoftaar T.M, Levy J.L, Salekshahrezaee Z and Vilanustree F, then the second cluster in green with two authors, consisting of Wang H and LV L.

Fig. 10. Screenshot of the VosViewer application related to the co-authoring network

In Figure 11, the first author cluster has khoshgoftaar t.m. most articles with a total of 4 articles, namely Salekshahrezaee [13], Leevy [14], Kennedy [15] and Wang & Khoshgoftaar [16], then the second largest in the second cluster is author Wang H with 2 articles, namely Du [17] and Wang & Khoshgoftaar [16].

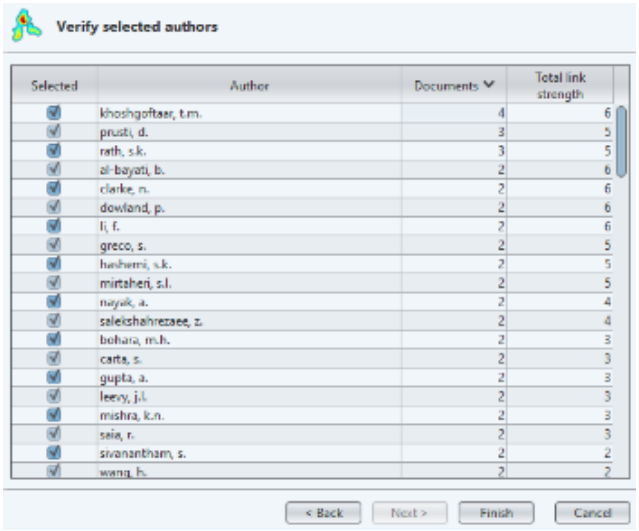

Fig. 11. Screenshot of the Vos Viewer application related to the co-authoring network.

F. *Influential and Trending Articles*

Table 1 shows the ranking of citations as indicated by the TC (Total Citations) value in the research topic.

TABLE I. RANKING OF ARTICLES BASED ON CITATIONS.

Id	Author	Title	Year	TC
1	Akbani [18]	Applying support vector machines to imbalanced datasets	2004	892
2	Randhawa [19]	Credit Card Fraud Detection Using AdaBoost and Majority Voting	2018	248
3	Lee [20]	Anomaly detection via online oversampling principal component analysis	2013	163
4	Taha & Malebary [11]	An Intelligent Approach to Credit Card Fraud Detection Using an Optimized Light Gradient Boosting Machine	2020	140
5	Carcillo [21]	SCARFF: A scalable framework for streaming credit card fraud detection with spark	2018	139

The article titled "Applying support vector machines to unbalanced datasets" with the author Akbani [18] has the first largest TC value with a value of 892, the article entitled "Credit Card Fraud Detection Using AdaBoost and Majority Voting" with the author Randhawa [19] has the second largest with a value TC value of 248, and the article entitled "Anomaly Detection via Online Oversampling Principal Component Analysis" with the author Lee [20] has the third largest TC value with a value of 163.

G. *Co-writing network*

The minimum co-occurrence scale for a keyword is 10. Of the 4854 keywords used in this research topic, only 85 met the threshold (criteria). Then from 85 selected keywords to 40 keywords. Of the 40 selected keywords, 60% were adopted, and 24 keywords were considered relevant to the research topic. Based on these 24 keywords, the Vos Viewer application is divided into three main clusters. Each colour represents a group with association links between keywords. The first large group is red, this group has 9 keywords, namely data mining, decision tree, fraud transaction, k nearest neighbor, logistic regression, random forest and support vector machine. The second largest group is coloured green, this group has 8 keywords, namely comparative analysis, deep learning, detection, effectiveness, fraudster, fraudulent credit card, machine learning mode and XG Boost, and the third largest group is coloured blue, this group has 7 keywords namely classifier, comparison, F1 score, fraudulent activity, machine learning, precision and recall. The three main clusters can be seen in Figure 12.

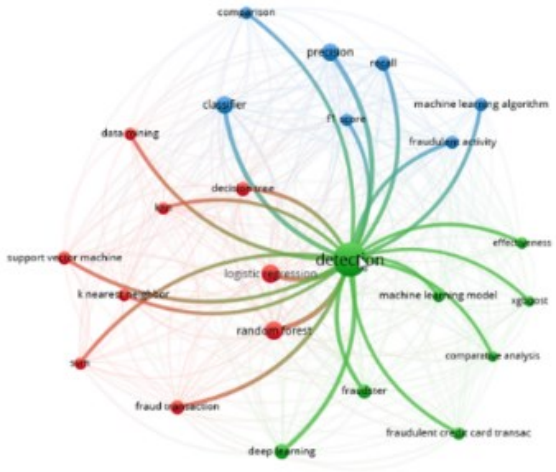

Fig. 12. Screenshot of the Vos Viewer keyword analysis application

In Figure 13, the time-based analysis of keywords is divided into three clusters. The first cluster is keywords that will be used in the future (2021.5). These keywords are XG Boost, machine learning model, fraudulent activity, precision, recall, F1 score, KNN and fraud transaction. This cluster is shown in yellow. The second cluster is keywords used today (2021.0). These keywords are detection, classifier, decision tree, machine learning algorithm, k nearest neighbour, logistic regression, random forest, comparative analysis, fraudulent credit card, deep learning, fraudster and comparison. This cluster is shown in green. The third cluster is keywords used in the past (2020.5). These keywords are effectiveness, data mining, support vector machine and SVM. This cluster is shown in blue.

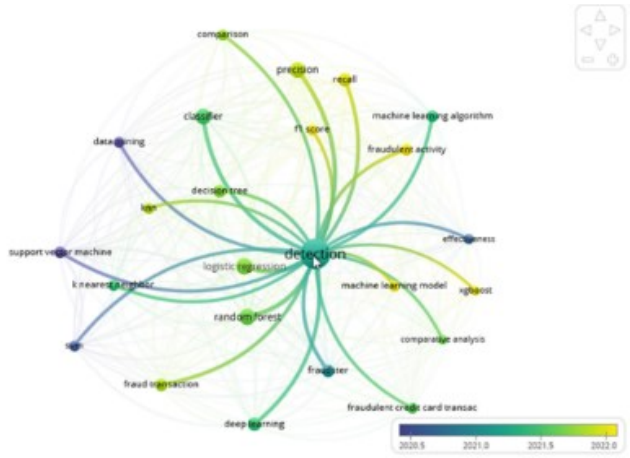

Fig. 13. Screenshot of the Vos Viewer keyword analysis application over time

IV. Discussion

The content analysis based on 40 keywords is divided into three research streams, which are interconnected in the research topic. The first is the research branch on the use of machine learning in data mining. This research stream is based on the keywords data mining, decision tree, fraud transaction, k nearest neighbour, logistic regression, random forest and support vector machine. Machine learning techniques include k nearest neighbours, logistic regression, random forest, decision trees, and support vector machines. This technique can detect irregular patterns that indicate fraudulent behaviour in credit card transactions [3] [5] [6] [22-23].

Decision trees are often used to detect fraudulent transactions, particularly in the context of credit card fraud. Researchers have used decision trees as part of intelligent approaches to credit card fraud detection, often combined with data mining [11]. The KNN (K-Nearest Neighbour) algorithm has been extensively studied in the context of fraudulent transaction detection, particularly in credit card fraud detection. KNN is a popular machine learning algorithm that classifies data points based on the majority class of their k nearest neighbours. In a comparative analysis of credit card fraud detection techniques, ANN was evaluated alongside machine learning and deep learning methods to determine the most efficient algorithm for detecting fraudulent transactions [5]. The use of Support Vector Machines (SVM) to detect fraudulent transactions has been the subject of extensive research and application. SVM has been recognized for its exceptional success in various applications including fraud detection by Akbani[18]. The use of logistic regression techniques to detect fraudulent transactions has become an interesting topic in the field of fraud detection. Logistic regression has proven to be an excellent tool for detecting financial fraud, particularly in the context of credit card transactions [22]. Random Forest can be used to overcome classification errors due to data imbalance, so Random Forest can be effective in building fraud detection models based on imbalanced data [22]. The use of multiple techniques in machine learning can lead to more efficient and effective accuracy of fraud detection models.

Second is the research focus on the use of deep learning. This research stream is based on the comparative analysis of the keyword, deep learning, detection, effectiveness, fraudster, fraudulent credit card, machine learning mode and XG Boost. Deep learning techniques are Deep Neural Networks (DNN) and Convolutional Neural Networks (CNN). These techniques have been used to overcome the challenges associated with detecting fraudulent credit card transactions. Researchers have proposed deep learning to develop online credit card fraud detection models using datasets such as Kaggle [6]. A comparative analysis was conducted to evaluate the effectiveness of various machine learning and deep learning techniques, including DNN and CNN, in detecting fraudulent transactions in credit card records [5]. Deep learning techniques have been integrated with resampling techniques to address the issue of dataset imbalance in credit card fraud detection. Efficient resampling methods combined with deep learning have been investigated to improve the detection of fraudulent credit card transactions [8]. Deep learning models, particularly those based on multi-layer perceptron (MLP) artificial neural networks, have been used to develop predictive analysis-based fraud detection systems for credit/debit card transactions [24]. The application of deep learning in credit card fraud detection has been compared with machine learning algorithms such as Random Forest and Support Vector Machine (SVM) to evaluate their effectiveness in detecting fraudulent activities [2]. The use of deep learning along with feature engineering and sampling strategies is used to overcome limited and inhomogeneous transaction data [25]. This approach holds promise for overcoming the complexity of credit card transaction data and improving the accuracy of fraud detection systems.

In third place is the research branch on machine learning for classifiers. This research stream is based on the classifier, comparison, F1 score, fraudulent activity, machine learning, precision and recall of the keyword. The impracticality of directly using historical transaction records to classify fraudulent credit card transactions based on the high volume and variety of transactions was highlighted by Moumeni[3]. Machine learning methods can detect significant credit card fraud cases [6]. The application of resampling and boosting classifier strategies has been studied to improve the effectiveness of machine learning algorithms in classifying anomalous patterns in financial transactions with an imbalanced class distribution [9] [22].

This research is limited by several factors. First, the processed articles only use Scopus Web as a database to extract relevant studies. This is due to inherent limitations of some bibliometric analysis and visualization software. Research on the research topic Machine learning for credit card fraud detection is an evolving area and there are several new studies available in databases other than the Scopus database, such as Web of Science, so it is overlooked. Future research can combine data from all databases for bibliometric analysis so that the research results are more comprehensive. Second, the limitations of the Vos Viewer and Scopus web application used for bibliometric analysis and visualization. If some of the newest and highest quality articles do not have many citations in citation analysis due to the latest publication date, Vos Viewer and Scopus web application will ignore them in citation analysis. In the meantime, expect the article to be widely cited in the future.

V. Conclusion

The main points of bibliometric analysis are three main research streams that are interconnected in research, namely: the research stream that uses machine learning in data mining, the research stream that uses deep learning for recognition, the research stream that uses machine learning for classifiers. The

number of publications related to research on the use of Machine Learning for credit card fraud detection is the highest in 2023 with 66 publications. The scientific article with the highest number of citations was 892 with the title "Applying support vector machines to imbalanced datasets". The author who has the most articles is the author Khoshgoftaar T.M with 4 articles, lovely professional university is the institution that produces the most articles with 5 articles. This paper contributes to the mapping process, provides potential reference information, and identifies gaps in existing research to advance machine learning research on credit card fraud detection in the future. Related up-to-date information for new professionals and researchers can be used to collaborate, test and improve ideas in the future. Future bibliometric analysis and visualization research on this topic could use the same methodology as the discussion of the current articles. Further research is recommended to include more specific keyword searches for better bibliometric analysis.

ACKNOWLEDGMENT

This work was supported by the research and community service institute (LP2M), UIN Maulana Malik Ibrahim Malang [Cost Standard Based Research: Grand Number 615, 2024].

REFERENCES

[1] S. Benson Edwin Raj and A. Annie Portia, "Analysis on credit card fraud detection methods," in *2011 International Conference on Computer, Communication and Electrical Technology (ICCCET)*, 2011, pp. 152–156. doi: 10.1109/ICCCET.2011.5762457.

[2] G. Sasikala et al., "An innovative sensing machine learning technique to detect credit card frauds in wireless communications," *Wireless Communications and Mobile Computing*, vol. 2022, 2022, doi: 10.1155/2022/2439205.

[3] L. Moumeni, M. Saber, I. Slimani, I. Elfarissi, and Z. Bougroun, "Machine learning for credit card fraud detection," in *Lecture Notes in Electrical Engineering, Bennani S., Lakhrissi Y., Khaissidi G., Mansouri A., and Khamlichi Y., Eds.*, 2022, pp. 211–221. doi: 10.1007/978-981-33-6893-4_20.

[4] A. Basit, M. Zafar, X. Liu, A. R. Javed, Z. Jalil, and K. Kifayat, "A comprehensive survey of AI-enabled phishing attacks detection techniques," *Telecommunication Systems*, vol. 76, no. 1, pp. 139–154, 2021, doi: 10.1007/s11235-020-00733-2.

[5] M. Ashraf, M. A. Abourezka, and F. A. Maghraby, "A comparative analysis of credit card fraud detection using machine learning and deep learning techniques," in *Lecture Notes in Networks and Systems, Magdi D.A., Helmy Y.K., Mamdouh M., and Joshi A., Eds*., 2022, pp. 267–282. doi: 10.1007/978-981-16-2275-5_16.

[6] A. Alharbi et al., "A novel text2IMG mechanism of credit card fraud detection: A deep learning approach," *Electronics*, vol. 11, no. 5, 2022, doi: 10.3390/electronics11050756.

[7] M. S. A. Alias, N. Ibrahim, and Z. M. Zin, "Comparative study of machine learning algorithms and correlation between input parameters," *International Journal of Integrated Engineering*, vol. 11, no. 4, pp. 81–90, 2019, doi: 10.30880/ijie.2019.11.04.009.

[8] P. Mrozek, J. Panneerselvam, and O. Bagdasar, "Efficient resampling for fraud detection during anonymised credit card transactions with unbalanced datasets," in *Proc. - IEEE/ACM Int. Conf. Util. Cloud Comput.*, 2020, pp. 426–433. doi: 10.1109/UCC48980.2020.00067.

[9] M. Valavan and S. Rita, "Predictive-analysis-based machine learning model for fraud detection with boosting classifiers," *Computer Systems Science and Engineering,* vol. 45, no. 1, pp. 231–245, 2023, doi: 10.32604/csse.2023.026508.

[10] D. Choi and K. Lee, "An artificial intelligence approach to financial fraud detection under IoT environment: A survey and implementation," *Security and Communication Networks*, vol. 2018, 2018, doi: 10.1155/2018/5483472.

[11] A. A. Taha and S. J. Malebary, "An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine," *IEEE Access*, vol. 8, pp. 25579–25587, 2020, doi: 10.1109/ACCESS.2020.2971354.

[12] A. H. Almuteer, A. A. Aloufi, W. O. Alrashidi, J. F. Alshobaili, and D. M. Ibrahim, "Detecting credit card fraud using machine learning," *International Journal of Interactive Mobile Technologies*, vol. 15, no. 24, pp. 108–122, 2021, doi: 10.3991/IJIM.V15I24.27355.

[13] Z. Salekshahrezaee, J. L. Leevy, and T. M. Khoshgoftaar, "A reconstruction error-based framework for label noise detection," *Journal of Big Data*, vol. 8, no. 1, 2021, doi: 10.1186/s40537-021-00447-5.

[14] J. L. Leevy, J. Hancock, T. M. Khoshgoftaar, and A. Abdollah Zadeh, "Investigating the effectiveness of one-class and binary classification for fraud detection," *Journal of Big Data*, vol. 10, no. 1, 2023, doi: 10.1186/s40537-023-00825-1.

[15] R. K. L. Kennedy, Z. Salekshahrezaee, F. Villanustre, and T. M. Khoshgoftaar, "Iterative cleaning and learning of big highly-imbalanced fraud data using unsupervised learning," *Journal of Big Data*, vol. 10, no. 1, 2023, doi: 10.1186/s40537-023-00750-3.

[16] H. Wang and T. M. Khoshgoftaar, "Review of feature selection techniques for credit card fraud detection," in *ISSAT Int. Conf. Reliab. Qual. Des., RQD*, 2023, pp. 396–401.

[17] D. Cheng, X. Wang, Y. Zhang, and L. Zhang, "Graph neural network for fraud detection via spatial-temporal attention," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 8, pp. 3800–3813, 2022, doi: 10.1109/TKDE.2020.3025588.

[18] R. Akbani, S. Kwek, and N. Japkowicz, "Applying support vector machines to imbalanced datasets," in *Lecture Notes in Computer Science*, *Boulicaut, J. F., Esposito, F., Giannotti, F., Pedreschi, D., Eds.*, 2004, pp. 39–50. doi: 10.1007/978-3-540-30115-8_7.

[19] K. Randhawa, C. K. Loo, M. Seera, C. P. Lim, and A. K. Nandi, "Credit card fraud detection using AdaBoost and majority voting," *IEEE Access*, vol. 6, pp. 14277–14284, 2018, doi: 10.1109/ACCESS.2018.2806420.

[20] Y.-J. Lee, Y.-R. Yeh, and Y.-C. F. Wang, "Anomaly detection via online oversampling principal component analysis," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 7, pp. 1460–1470, 2013, doi: 10.1109/TKDE.2012.99.

[21] F. Carcillo, A. Dal Pozzolo, Y.-A. Le Borgne, O. Caelen, Y. Mazzer, and G. Bontempi, "SCARFF: A scalable framework for streaming credit card fraud detection with spark," *Information Fusion*, vol. 41, pp. 182–194, 2018, doi: 10.1016/j.inffus.2017.09.005.

[22] N. M. Mqadi, N. Naicker, and T. Adeliyi, "Solving misclassification of the credit card imbalance problem using near miss," *Mathematical Problems in Engineering*, vol. 2021, 2021, doi: 10.1155/2021/7194728.

[23] S. K. Hashemi, S. L. Mirtaheri, and S. Greco, "Fraud detection in banking data by machine learning techniques," *IEEE Access*, vol. 11, pp. 3034–3043, 2023, doi: 10.1109/ACCESS.2022.3232287.

[24] B. Kasasbeh, B. Aldabaybah, and H. Ahmad, "Multilayer perceptron artificial neural networks-based model for credit card fraud detection," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 26, no. 1, pp. 362–373, 2022, doi: 10.11591/ijeecs.v26.i1.pp362-373.

[25] Y.-Y. Hsin, T.-S. Dai, Y.-W. Ti, M.-C. Huang, T.-H. Chiang, and L.-C. Liu, "Feature engineering and resampling strategies for fund transfer fraud with limited transaction data and a time-inhomogeneous modi operandi," *IEEE Access*, vol. 10, pp. 86101–86116, 2022, doi: 10.1109/ACCESS.2022.3199