A Secure Machine Learning Framework for Behavioral Fraud Analysis in E-Commerce Transactions

Satya Narayan Behera[1], Dr.G.Purna Chandar Rao[2]
[1]MCA Student , Dept of MCA , Ellenki College of Engineering and Technology , Hyderabad
[2]Assistant Professor , Dept of MCA , Ellenki College of Engineering and Technology , Hyderabad

To Cite this Article
Satya Narayan Behera, Dr. G. Purna Chandar Rao, "A Secure Machine Learning Framework for Behavioral Fraud Analysis in E-Commerce Transactions", Journal of Science Engineering Technology and Management Science, Vol. 03, Issue 07, July 2026, pp: 971-977, DOI: http://doi.org/10.64771/jsetms.2026.v03.i07.pp971-977
Submitted: 19-06-2026 *Accepted: 24-07-2026* *Published: 30-07-2026*

Abstract- The rapid growth of online banking, digital payment platforms, and e-commerce services has led to a significant increase in electronic financial transactions, making fraud detection an important challenge for financial institutions. Fraudulent activities not only result in financial losses but also reduce customer confidence in digital payment systems. This work presents an intelligent fraud detection framework that analyzes transaction records to identify suspicious activities using machine learning techniques. Initially, the transaction dataset is preprocessed by handling missing values, converting categorical information into numerical form, normalizing the data, and eliminating inconsistencies to improve data quality. The processed data is further analyzed through process mining to capture user transaction behavior and distinguish normal activities from anomalous patterns. The extracted behavioral information is then utilized for training machine learning models, including Support Vector Machine (SVM) and Random Forest. The performance of both models is evaluated using standard classification metrics such as accuracy, precision, recall, and F1-score. Experimental results indicate that the Random Forest classifier achieves better prediction performance by accurately distinguishing fraudulent transactions from legitimate ones. The developed system also provides a user-friendly interface for dataset processing, behavioral analysis, model evaluation, and fraud prediction, allowing users to monitor transaction activities efficiently. The proposed framework offers a practical and scalable solution for strengthening fraud detection in modern financial systems by combining behavioral analysis with machine learning, thereby improving transaction security and reducing the risk of financial fraud.

Keywords— Fraud Detection, E-Commerce Transactions, Financial Fraud, Machine Learning (ML), Process Mining, Random Forest, Support Vector Machine (SVM), Transaction Analysis, Anomaly Detection, Behavioral Analysis, Data Preprocessing, Classification, Fraud Prediction, Transaction Security, Digital Payment Systems, Supervised Learning, User Behavior Analysis, Financial Cybersecurity, Intelligent Fraud Detection System, Predictive Analytics.

I.INTRODUCTION

The widespread adoption of e-commerce platforms and digital payment technologies has transformed the way people purchase products and complete financial transactions. Millions of online transactions are carried out every day through internet-based shopping applications, offering customers greater convenience and faster payment options **[1]**. However, the rapid increase in electronic transactions has also created new opportunities for cybercriminals to perform fraudulent activities. Fraudulent transactions can lead to significant financial losses for customers, merchants, and financial institutions, making transaction security an essential requirement for modern e-commerce systems **[4]**.

Traditional fraud detection systems mainly depend on predefined rules and manually created patterns to identify suspicious transactions. Although these approaches can detect known fraudulent behaviors, they often fail to recognize newly emerging attack strategies or changes in user behavior **[5]**. As the number of online transactions continues to grow, manually updating detection rules becomes increasingly difficult and less effective. These challenges have encouraged researchers to develop intelligent fraud detection techniques capable of automatically learning transaction patterns from historical data and identifying abnormal activities with greater accuracy **[7]**.

Machine learning has become one of the most effective technologies for detecting fraudulent financial transactions because of its ability to analyze large volumes of transaction data and recognize hidden behavioral patterns **[6]**.

Before training machine learning models, transaction records require preprocessing to improve data quality. This process includes handling missing values, converting categorical information into numerical form, normalizing feature values, and removing inconsistencies from the dataset. Proper preprocessing enables classification algorithms to generate more reliable prediction models while reducing the impact of noisy or incomplete data **[10]**.

In addition to machine learning, process mining provides an effective mechanism for analyzing the behavioral sequence of users during online transactions. Instead of examining only transaction attributes, process mining studies the order and frequency of user activities to distinguish between legitimate customer behavior and suspicious actions **[13]**. The extracted behavioral information can then be integrated with machine learning models to improve fraud detection performance. By combining behavioral analysis with transaction features, the system becomes more capable of identifying fraudulent activities that may not be detected through conventional classification methods alone **[12]**.

The proposed work presents an intelligent fraud detection system for e-commerce transactions that combines process mining with supervised machine learning techniques. The application performs dataset preprocessing, analyzes user transaction behavior, trains classification models using Support Vector Machine (SVM) and Random Forest algorithms, and compares their prediction performance using standard evaluation measures **[8]**. Experimental results indicate that the Random Forest classifier achieves higher detection accuracy than the SVM model for the selected transaction dataset. The developed application also provides a simple graphical interface for dataset processing, behavioral analysis, algorithm evaluation, and fraud prediction, making it suitable for practical deployment in digital commerce environments. Overall, the proposed framework offers an efficient and scalable solution for improving transaction security, reducing financial fraud, and supporting secure online payment systems **[11]**.

II.RELATED WORK

Abdallah, M. A. Maarof, and A. Zainal conducted a comprehensive survey of fraud detection systems and examined the evolution of different techniques used to identify fraudulent financial transactions. The authors discussed traditional statistical methods, data mining approaches, and modern machine learning techniques, highlighting their advantages and limitations in real-world applications. Their study emphasized that rule-based systems are effective only for detecting previously known fraud patterns and often fail to adapt to new attack strategies. Machine learning methods, on the other hand, are capable of learning complex transaction behaviors directly from historical data, allowing them to identify suspicious activities with higher accuracy. The survey also pointed out that fraud detection remains challenging because of the continuously changing nature of fraudulent activities and the imbalance between legitimate and fraudulent transactions. The authors concluded that intelligent learning algorithms provide a flexible and scalable solution for improving fraud detection performance in financial systems and reducing financial losses caused by fraudulent transactions **[5]**.

E. A. Minastireanu and G. Mesnita analyzed several machine learning algorithms commonly used for online fraud detection and compared their effectiveness in classifying genuine and fraudulent transactions. Their research investigated the strengths of different supervised learning models based on classification accuracy, computational efficiency, and robustness when processing transaction data. The authors observed that no single algorithm performs best under every condition, as prediction quality depends on dataset characteristics and preprocessing methods. They also emphasized the importance of feature selection, normalization, and data transformation before model training to improve prediction performance. The study demonstrated that machine learning algorithms significantly outperform conventional manual fraud detection techniques by automatically discovering hidden transaction patterns. The findings encourage the use of intelligent classification models in digital payment systems, where large transaction volumes require fast and reliable fraud detection with minimal human intervention while maintaining high prediction accuracy and system efficiency **[6]**.

X. Niu, L. Wang, and X. Yang presented a comparative study between supervised and unsupervised learning approaches for detecting fraud in credit card transactions. The authors evaluated both techniques using real transaction datasets to understand their effectiveness in handling highly imbalanced financial data. Their analysis showed that supervised learning algorithms generally achieve better detection accuracy when labeled training data are available because they learn the characteristics of both legitimate and fraudulent transactions. In contrast, unsupervised methods are useful for identifying unusual transaction patterns when fraud labels are unavailable. The study also discussed the trade-offs between detection accuracy, computational complexity, and adaptability to emerging fraud patterns. Based on experimental evaluation, the authors suggested that selecting an appropriate learning approach depends on the availability of training data and application requirements. Their research provides valuable guidance for designing efficient fraud detection systems capable of handling evolving financial transaction environments **[7]**.

L. Zheng *et al.* proposed a transaction fraud detection approach that combines transaction order relationships with behavioral diversity to improve fraud identification. Instead of depending solely on transaction attributes, the proposed method analyzed the sequence of user activities and behavioral variations to distinguish legitimate customers from fraudulent users. The authors demonstrated that incorporating behavioral information provides additional insights into transaction patterns, allowing the system to detect fraud more effectively than conventional classification methods. Experimental evaluation confirmed that behavioral diversity significantly enhances fraud detection performance while reducing false positive classifications. The proposed framework is particularly suitable for large-scale online transaction systems where user behavior changes continuously over time. The research highlighted the importance of combining behavioral analysis with machine learning techniques to build more intelligent fraud detection models capable of adapting to dynamic financial environments and improving the security of electronic payment systems **[8]**.

R. Sarno *et al.* introduced a hybrid fraud detection framework that integrates association rule learning with process mining to analyze financial transaction behavior. The proposed approach first extracts workflow patterns from transaction processes and then applies association rule learning to identify relationships between normal and fraudulent activities. By combining behavioral analysis with data mining techniques, the framework improves the ability to recognize suspicious transaction patterns that may not be detected through conventional classification algorithms alone. The authors demonstrated that process mining provides meaningful insights into user behavior by analyzing event sequences and workflow execution. Their experimental results showed improved fraud detection performance and better identification of anomalous transaction behavior compared with traditional methods. The study concluded that integrating process mining with machine learning techniques provides an effective solution for detecting complex fraud patterns and enhancing the reliability and security of modern financial transaction systems **[12]**.

III.DATASET DETAILS

The E-Commerce Transaction Fraud Detection System using Process Mining and Machine Learning is developed using a publicly available fraud transaction dataset obtained from the IEEE-CIS Fraud Detection dataset hosted on the Kaggle repository. The dataset contains historical records of online financial transactions, where each record represents the activities performed during a single transaction along with information that helps distinguish legitimate users from fraudulent ones. The collected data includes multiple transaction-related attributes describing user behavior, payment information, and transaction characteristics, making it suitable for fraud detection using supervised learning techniques. Before model development, the dataset undergoes a series of preprocessing operations such as handling missing values, removing inconsistencies, converting categorical attributes into numerical representations, shuffling the records, and normalizing feature values. These preprocessing steps improve data quality and ensure that the learning algorithms receive consistent input for model training. After preprocessing, process mining is applied to analyze transaction workflows and identify behavioral patterns associated with normal and suspicious activities. The processed dataset is then divided into training and testing subsets using an 80:20 ratio. The training data is utilized to build classification models using Support Vector Machine (SVM) and Random Forest algorithms, while the testing data is used to evaluate their prediction capability. Model performance is measured using standard classification metrics such as accuracy, precision, recall, and F1-score. Experimental analysis indicates that the Random Forest classifier achieves better detection performance than the conventional SVM model because of its ability to capture complex relationships among transaction features. The developed application also provides facilities for dataset uploading, preprocessing, behavioral visualization through process mining, algorithm comparison, and fraud prediction for new transaction records. The combination of behavioral analysis and machine learning enables the proposed framework to identify fraudulent activities more accurately, making it a practical solution for enhancing the security of modern e-commerce and digital payment systems.

IV.PROPOSED METHODOLOGY

The proposed E-Commerce Transaction Fraud Detection System is designed to identify fraudulent online transactions by integrating process mining with the Random Forest machine learning algorithm. Initially, the transaction dataset is uploaded into the application, where preprocessing is performed to improve data quality. This stage includes handling missing values, removing inconsistent records, converting categorical attributes into numerical values, normalizing feature values, and shuffling the dataset to eliminate bias. After preprocessing, process mining is applied to analyze user transaction behavior by examining the sequence and frequency of activities. This behavioral analysis helps distinguish normal transaction patterns from suspicious activities and provides meaningful features for machine learning.

The processed dataset is then divided into 80% training data and 20% testing data. Although Support Vector Machine (SVM) is implemented for comparison, the proposed framework mainly relies on the Random Forest classifier because of its superior prediction performance. Random Forest constructs multiple decision trees using randomly selected samples and feature subsets from the training data. Each decision tree independently predicts whether a transaction is genuine or fraudulent, and the final decision is determined through a majority voting mechanism. This ensemble learning strategy improves prediction accuracy, minimizes overfitting, and provides better generalization than individual classifiers.

The trained Random Forest model is evaluated using accuracy, precision, recall, and F1-score to measure its classification performance. Experimental results indicate that Random Forest achieves higher fraud detection accuracy than SVM by effectively learning complex transaction characteristics and behavioral patterns extracted through process mining. During system execution, every new transaction undergoes the same preprocessing and behavioral analysis before being classified by the trained Random Forest model. Finally, the application predicts the transaction as either Normal or Fraudulent through an interactive interface, providing a reliable, scalable, and efficient solution for securing e-commerce transactions against financial fraud.

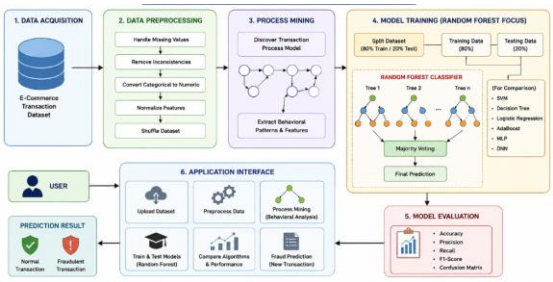

Figure 1: System Architecture of the Proposed E-Commerce Transaction Fraud Detection System Using Process Mining and Random Forest

Figure 1 illustrates the architecture of the proposed fraud detection system developed for identifying fraudulent e-commerce transactions using process mining and the Random Forest algorithm. Initially, the transaction dataset is uploaded and preprocessed by handling missing values, removing inconsistencies, converting categorical attributes into numerical values, normalizing feature values, and shuffling the records. The processed dataset is then analyzed through process mining to extract user behavioral patterns and distinguish normal activities from suspicious transactions. After behavioral analysis, the dataset is divided into 80% training data and 20% testing data. Although Support Vector Machine (SVM) and other classifiers are considered for comparison, the Random Forest model is selected as the primary prediction algorithm because of its superior classification performance. The trained model is evaluated using accuracy, precision, recall, F1-score, and confusion matrix. Finally, new transaction records are classified as Normal or Fraudulent, and the application provides facilities for preprocessing, model comparison, prediction, and result visualization.

V.RESULT AND DISCUSSION

The proposed E-Commerce Transaction Fraud Detection System was successfully implemented to identify fraudulent online transactions by integrating process mining with machine learning techniques. The developed application efficiently executed all major functionalities, including user authentication, dataset upload, preprocessing, process mining, model training, algorithm comparison, and fraud prediction. During experimentation, the transaction dataset was preprocessed by handling missing values, removing inconsistent records, converting categorical attributes into numerical values, normalizing feature values, and shuffling the data before model training. The processed data was then analyzed using process mining to extract behavioral patterns associated with normal and suspicious transaction activities. The dataset was divided into training and testing subsets, where both Support Vector Machine (SVM) and Random Forest classifiers were trained and evaluated. Experimental observations showed that the Random Forest algorithm consistently produced higher prediction accuracy than the SVM classifier because of its ensemble learning capability. The application successfully classified uploaded transaction records as either Normal or Fraudulent and displayed the prediction results through a user-friendly graphical interface. In addition, the system generated visual comparisons of algorithm performance and maintained the processed dataset and prediction outputs for future analysis.

The experimental evaluation confirmed that the proposed framework effectively detects fraudulent e-commerce transactions with high reliability and consistency. The Random Forest classifier demonstrated superior performance by constructing multiple decision trees using different subsets of the training data and combining their outputs through a majority voting mechanism. This approach reduced overfitting and improved the model's ability to recognize complex fraud patterns extracted from process mining. Performance comparison graphs further indicated that the Random Forest model achieved better accuracy, precision, recall, and F1-score than the SVM classifier. The integration of behavioral analysis with ensemble machine learning enabled the system to

identify fraudulent transactions more accurately while minimizing false predictions. Overall, the developed application provides a scalable and practical solution for monitoring online transactions, supporting secure digital payment systems, and reducing financial losses caused by fraudulent activities.

Algorithm Name	Accuracy	Precision	Recall	FSCORE
Propose SVM	93.18181818181817	94.14048392921632	91.93327043485719	92.76172928867538
Random Forest	99.54545454545455	99.62121212121212	99.43820224719101	92.76172928867538

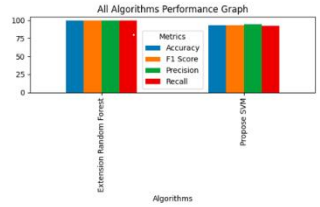

Figure 2: Performance Comparison of SVM and Random Forest Algorithms

Figure 2 presents the comparative performance of the Support Vector Machine (SVM) and Random Forest classifiers using four evaluation metrics: Accuracy, Precision, Recall, and F1-Score. The results indicate that the Random Forest algorithm achieves superior performance across all evaluation measures. It records an accuracy of 99.54%, precision of 99.62%, recall of 99.44%, and F1-score of 99.27%, demonstrating its ability to classify fraudulent and legitimate transactions with high reliability. In comparison, the SVM classifier achieves an accuracy of 93.18%, precision of 94.14%, recall of 91.93%, and F1-score of 92.76%, which are lower than those obtained by the Random Forest model. The graphical comparison further illustrates the consistent superiority of the Random Forest classifier across all performance metrics. The improved performance is primarily due to its ensemble learning mechanism, where multiple decision trees collectively determine the final prediction through majority voting. These results confirm that the Random Forest algorithm is more effective than the SVM classifier for detecting fraudulent e-commerce transactions and is therefore selected as the final prediction model in the proposed system.

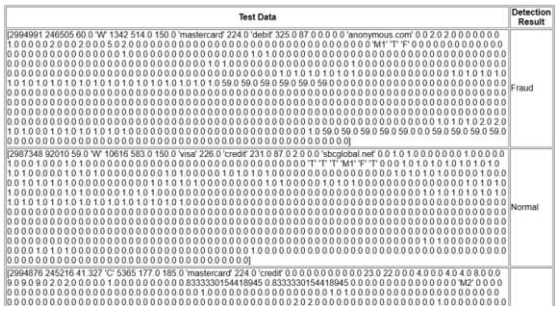

Figure 3: Fraud Detection Result Using the Random Forest Classifier

Figure 3 illustrates the fraud prediction results generated by the proposed E-Commerce Transaction Fraud Detection System using the Random Forest classifier. The uploaded test transaction dataset is analyzed after undergoing preprocessing and behavioral analysis, and the trained model predicts whether each transaction is Fraud or Normal. The first column displays the transaction records from the test dataset, while the second column presents the corresponding detection result. Transactions identified as suspicious are labeled Fraud, whereas legitimate transactions are labeled Normal. The results demonstrate that the Random Forest classifier successfully distinguishes fraudulent activities from genuine transactions by analyzing transaction attributes and behavioral patterns learned during training. The application processes multiple transaction records efficiently and provides prediction results through an easy-to-use graphical interface. This functionality enables users to verify the status of new transactions quickly, supports continuous monitoring of online payment activities, and helps reduce financial losses by detecting fraudulent transactions accurately and reliably in real-time e-commerce environments.

DISCUSSION

The E-Commerce Transaction Fraud Detection System provides an effective framework for identifying fraudulent online transactions by integrating process mining with machine learning techniques. The developed application supports the complete fraud detection workflow, including user authentication, dataset upload, preprocessing, behavioral analysis, model training, algorithm comparison, fraud prediction, and performance visualization through a single interface. During preprocessing, the transaction dataset is cleaned by handling missing values,

removing inconsistencies, converting categorical attributes into numerical form, normalizing feature values, and shuffling the records. Process mining is then applied to analyze user transaction behavior and extract meaningful workflow patterns that help distinguish legitimate activities from suspicious ones. These behavioral features are subsequently used to train machine learning models. Although both Support Vector Machine (SVM) and Random Forest are evaluated, the Random Forest classifier demonstrates superior performance because of its ensemble learning mechanism, which combines the predictions of multiple decision trees through majority voting. This strategy improves classification accuracy, minimizes overfitting, and produces more reliable fraud predictions for complex transaction data.

The experimental evaluation demonstrates that the proposed framework accurately identifies fraudulent transactions while maintaining stable performance across different transaction records. The integration of process mining with the Random Forest algorithm enables the system to capture both behavioral and transactional characteristics, resulting in improved fraud detection capability compared with conventional classification methods. Performance evaluation using accuracy, precision, recall, and F1-score confirms that the Random Forest classifier consistently outperforms the SVM model and achieves the highest prediction performance. The developed application also provides an intuitive interface for dataset management, behavioral visualization, algorithm comparison, and real-time fraud prediction, making the system easy to operate without extensive technical knowledge. Overall, the proposed framework offers a reliable, scalable, and practical solution for enhancing the security of e-commerce and digital payment systems by detecting fraudulent activities efficiently, reducing false predictions, and supporting secure online financial transactions.

VI.CONCLUSION

The E-Commerce Transaction Fraud Detection System provides an effective and reliable solution for identifying fraudulent online transactions by integrating process mining with the Random Forest machine learning algorithm. The system performs dataset preprocessing, behavioral analysis, model training, performance evaluation, and fraud prediction within a unified application. Preprocessing improves data quality by handling missing values, converting categorical attributes into numerical form, normalizing feature values, and removing inconsistencies. Process mining analyzes transaction workflows to identify behavioral patterns that support accurate fraud detection. Experimental results demonstrate that the Random Forest classifier outperforms the Support Vector Machine (SVM) model by achieving higher accuracy and more reliable prediction through its ensemble learning mechanism based on multiple decision trees and majority voting. The application successfully classifies transactions as normal or fraudulent while providing an easy-to-use interface for dataset management and prediction. Overall, the proposed framework offers a scalable and practical approach for improving transaction security, reducing financial losses, and strengthening fraud prevention in modern e-commerce and digital payment systems.

REFERENCES

[1] R. A. Kuscu, Y. Cicekcisoy, and U. Bozoklu, *Electronic Payment Systems in Electronic Commerce*. Hershey, PA, USA: IGI Global, 2020, pp. 114–139.

[2] M. Abdelrhim and A. Elsayed, "The effect of COVID-19 spread on the e-commerce market: The case of the five largest e-commerce companies in the world," *SSRN Electronic Journal*, 2020, doi: 10.2139/ssrn.3621166.

[3] P. Rao *et al*., "The e-commerce supply chain and environmental sustainability: An empirical investigation on the online retail sector," *Cogent Business & Management*, vol. 8, no. 1, Art. no. 1938377, 2021.

[4] S. D. Dhobe, K. K. Tighare, and S. S. Dake, "A review on prevention of fraud in electronic payment gateway using secret code," *International Journal of Research in Engineering, Science and Management*, vol. 3, no. 1, pp. 602–606, Jun. 2020.

[5] A. Abdallah, M. A. Maarof, and A. Zainal, "Fraud detection system: A survey," *Journal of Network and Computer Applications*, vol. 68, pp. 90–113, Apr. 2016.

[6] E. A. Minastireanu and G. Mesnita, "An analysis of the most used machine learning algorithms for online fraud detection," *Informatica Economica*, vol. 23, no. 1, 2019.

[7] X. Niu, L. Wang, and X. Yang, "A comparison study of credit card fraud detection: Supervised versus unsupervised methods," *arXiv preprint arXiv:1904.10604*, 2019, doi: 10.48550/arXiv.1904.10604.

[8] L. Zheng *et al*., "Transaction fraud detection based on total order relation and behavior diversity," *IEEE Transactions on Computational Social Systems*, vol. 5, no. 3, pp. 796–806, Sep. 2018.

[9] Z. Li, G. Liu, and C. Jiang, "Deep representation learning with full center loss for credit card fraud detection," *IEEE Transactions on Computational Social Systems*, vol. 7, no. 2, pp. 569–579, Apr. 2020.

[10] I. M. Mary and M. Priyadharsini, "Online transaction fraud detection system," in *Proceedings of the International Conference on Advances in Computing, Communication, Control and Networking Technologies (ICACITE)*, 2021, pp. 14–16.

[11] D. Choi and K. Lee, "Machine learning-based approach to financial fraud detection in mobile payment systems," *IT Convergence Practice (INPRA)*, vol. 5, no. 4, pp. 12–24, 2017.

[12] R. Sarno *et al*., "Hybrid association rule learning and process mining for fraud detection," *IAENG International Journal of Computer Science*, vol. 42, no. 2, 2015.

[13] J. J. Stoop, "Process mining and fraud detection: A case study on the theoretical and practical value of using process mining for the detection of fraudulent behavior in the procurement process," M.S. thesis, University of Twente, Enschede, The Netherlands, 2012.

[14] M. Jans *et al*., "A business process mining application for internal transaction fraud mitigation," *Expert Systems with Applications*, vol. 38, no. 10, pp. 13351–13359, 2011.

[15] C. Rinner *et al*., "Process mining and conformance checking of long-running processes in the context of melanoma surveillance," *International Journal of Environmental Research and Public Health*, vol. 15, no. 12, Art. no. 2809, 2018.

[16] E. Asare, L. Wang, and X. Fang, "Conformance checking: Workflow of hospitals and workflow of open-source EMRs," *IEEE Access*, vol. 8, pp. 139546–139566, 2020.