Enhancing Security in Cross-Border IoT Transactions through Predictive Machine Learning using Bi-GRU

R.Lingeswari[1], S.Brindha[2]
1Ph.D Research Scholar, Department of Computer Science
2 Associate Professor, Department of Computer Science and Applications
1&2 St.Peter's Institute of Higher Education & Research, Chennai, Tamilnadu, India.
lavalingu@gmail.com. brindhas.mca@spiher.ac.in

Abstract:

The proliferation of Internet of Things (IoT) devices has ushered in a new era of connectivity, transforming various sectors within Industry 4.0. However, this surge in IoT adoption has also brought about challenges, particularly in securing cross-border transactions. Conventional approaches to vulnerability analysis rely on rule-based decision-making embedded within models, often failing to detect fraudulent activities in high-frequency cross-border transactions. In this paper, we propose an innovative predictive big data analytics framework that effectively addresses vulnerabilities in cross-border IoT transactions. By harnessing the power of machine learning, we delve into historical transaction data, thereby enabling financial institutions to mitigate risks associated with cross-border operations. Our methodology leverages a Bidirectional Gated Recurrent Unit (Bi-GRU) model to discern subtle patterns indicative of fraudulent behavior, significantly enhancing vulnerability assessment. Simulation results underscore the superiority of our approach compared to existing methods in identifying and combatting transactional vulnerabilities.

Keywords:

Internet of Things, Cross-Border Transactions, Industry 4.0, Predictive Machine Learning, Bidirectional Gated Recurrent Unit, Fraud Detection

1. Introduction

The Internet of Things (IoT) has sparked a revolutionary wave of interconnected devices, redefining the landscape of industries and paving the way for the Industry 4.0 ecosystem [1]. This transformation has introduced unprecedented levels of automation, data exchange, and efficiency across various sectors [2]. However, as IoT technologies continue to gain momentum, the vulnerabilities associated with their integration into cross-border transactions have become a pressing concern [3].

Vulnerability analysis has been the cornerstone of identifying weaknesses in systems, often relying on predefined rules embedded within models [4]. These models, though effective in certain contexts, fall short when it comes to detecting intricate fraudulent behaviors exhibited in the realm of high-frequency cross-border transactions [5]. The dynamic and rapidly evolving nature of these transactions can obscure patterns that would otherwise raise red flags in conventional analyses [6].

In response to these challenges, this paper presents an innovative approach that harnesses the power of predictive big data analytics and machine learning to fortify the security of cross-border transactions [7]. By delving into historical data logs from these transactions [8]-[10], financial institutions can proactively identify vulnerabilities and mitigate potential risks. Our proposed methodology hinges on the utilization of a Bidirectional Gated Recurrent Unit (Bi-GRU) model, a deep learning architecture designed to capture complex temporal patterns within sequences of data. This model enables us to discern nuanced patterns indicative of fraudulent activities, facilitating a comprehensive vulnerability assessment.

The research aim to bridge the gap between traditional rule-based vulnerability analysis and the dynamic nature of modern cross-border IoT transactions. By training our Bi-GRU model on historical data, we empower financial institutions to anticipate and thwart fraudulent activities, thereby minimizing financial losses and preserving the integrity of cross-border transactions. In the subsequent sections, we elaborate on the specifics of our approach, detailing the implementation of the Bi-GRU model and the methodology for predictive learning from historical data logs. We then present the results of our simulations, demonstrating the superior efficacy of our proposed method in comparison to existing techniques.

This paper presents a proactive strategy for tackling vulnerabilities within cross-border IoT transactions, leveraging predictive machine learning to enhance security and mitigate risks. By fostering a deeper understanding of transactional patterns and behaviors, our approach aims to bolster the resilience of financial systems operating in the rapidly evolving landscape of IoT-driven cross-border operations.

2. Related works

Huang et al. [11]]conduct an in-depth analysis of the utilization of deep learning models within the crucial domains of finance and banking. Their study systematically evaluates the preprocessing techniques, input data, and model assessment methods employed in these applications. Additionally, they explore three critical factors that could influence the outcomes of deep learning models in the context of financial tasks. This research contributes valuable insights and guidance for both academic researchers and industry practitioners, offering a comprehensive overview of the contemporary application of deep learning models within the domains of finance and banking.

Widyastuti et al. [12] present a study centered on classifying the quality of customer service at the Bank BTN Pematangsiantar Branch. Their dataset comprises customer questionnaire responses from the said bank's customers. The study employs seven attributes, including Age, Job, Old to Customer, Tangible, Reliability, Assurance, and Responsiveness. Utilizing the C4.5 Algorithm in conjunction with the RapidMiner software to develop decision trees, the study yields five classification rules for determining customer service quality. These rules are distributed across satisfaction and dissatisfaction statuses. Notably, the C4.5 algorithm attains an accuracy of 77.78% in assessing the quality of customer service at the Bank BTN Pematangsiantar Branch. This analysis is poised to aid the bank in enhancing service quality and subsequently elevating customer satisfaction levels.

Leo et al. [13] undertake an extensive review to explore machine-learning techniques applied to risk management within the banking sector. Their objective is to identify areas within risk management that remain underexplored in terms of machine learning application. While previous research has delved into the application of machine learning to manage banking risks such as credit, market, operational, and liquidity risks, the authors note a significant gap between the potential of machine learning and its current industry-level focus. There are substantial opportunities for further research to explore the application of machine learning to address specific challenges within bank risk management.

Karvana et al. [14] present a method developed for the analysis and prediction of customer churn. In their study, five distinct classification methods are tested using a dataset encompassing 57 attributes. Through multiple experimental iterations and class comparisons, the Support Vector

Machine (SVM) with a balanced 50:50 class sampling emerges as the optimal method for predicting customer churn at a private bank in Indonesia. The findings of this modeling exercise can be harnessed by organizations seeking to implement strategic measures to mitigate customer churn effectively.

3. Proposed Method

The proposed methodology leverages the predictive machine learning, specifically employing a Bidirectional Gated Recurrent Unit (Bi-GRU) model, to enhance security and mitigate risks associated with vulnerabilities in cross-border IoT transactions. The methodology comprises several interconnected stages, including dataset collection, pre-processing, feature extraction, and risk classification which is shown in Figure 1.

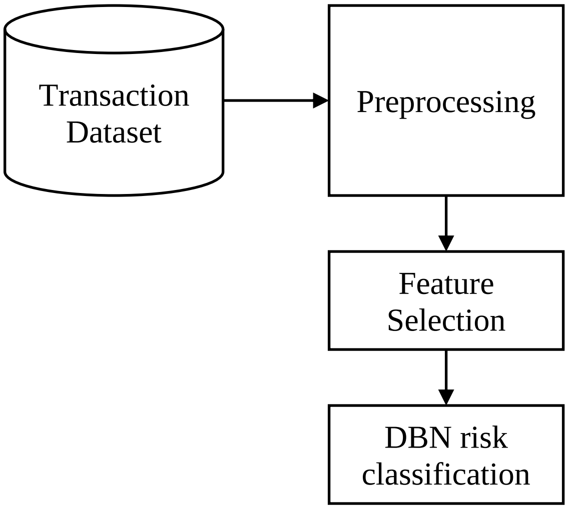

Figure 1: Proposed Model

3.1. Dataset Collection

The proposed approach lies in the availability of a comprehensive and representative dataset. This dataset is composed of historical cross-border IoT transactions, encompassing both legitimate and fraudulent cases. Collaborations with financial institutions or relevant regulatory bodies facilitate the acquisition of real-world transactional data. The dataset should encompass a diverse range of transaction types, geographic regions, and transaction volumes to ensure robust model training.

The consortium collaborates with various banks, payment processors, and financial technology companies across different regions. These entities contribute anonymized transactional data from their cross-border operations. The dataset includes a wide range of attributes, such as transaction timestamps, transaction amounts, originating and receiving countries, device information, and more.

The contributed datasets are collected and integrated into a centralized repository. The datasets may vary in structure, format, and level of detail due to differences in data collection practices among institutions. Data engineers work on standardizing the data format, resolving inconsistencies, and creating a unified schema to ensure that the dataset is ready for analysis.

To prepare the data for machine learning, the consortiums data science team works on labeling each transaction as either legitimate or fraudulent. Historical records of confirmed fraudulent transactions are collected from past investigations, while legitimate transactions are those with no evidence of fraud. These labels provide the ground truth for model training and evaluation.

The dataset is intentionally designed to encompass a wide variety of cross-border transaction scenarios. This includes transactions of varying magnitudes, transaction frequencies, regions, types of IoT devices involved, and types of goods or services exchanged. This diversity ensures that the model can generalize well and effectively detect vulnerabilities across different transaction contexts. The sensitive nature of financial data, data privacy and regulatory compliance are paramount. Personal information is anonymized and any potential identifiers are removed to ensure the privacy of individuals involved. The data collection process adheres to data protection regulations such as GDPR and follows industry best practices for data security. The collected data serves as the basis for training and evaluating the Bidirectional Gated Recurrent Unit (Bi-GRU) model, allowing financial institutions to proactively identify vulnerabilities and mitigate risks in cross-border IoT transactions.

3.2. Pre-processing

Preprocessing the dataset involves a series of data transformation steps to ensure that the data is clean, standardized, and ready for analysis. These steps are crucial to preparing the dataset for feeding into machine learning models.

Data Cleaning and Handling Missing Values: In the dataset, there might be missing values for certain attributes. One common approach to handle missing values is imputation, where missing values are replaced with estimated or calculated values.

Normalization: Normalization ensures that attributes with different scales do not disproportionately influence the model. The formula for min-max normalization is:

$$||x|| = (x - x_{min})/(x_{max} - x_{min})$$

Categorical Variable Encoding: Many attributes, such as transaction types (type) or countries (country), might be categorical. These need to be encoded numerically for machine learning algorithms. One-hot encoding is a common method, where each category becomes a binary column. For example, if type can be purchase, transfer, or withdrawal, one-hot encoding would create three binary columns.

Timestamp Alignment and Feature Engineering: If the dataset includes timestamps (timestamp), the pre-processor extract useful features like day of the week, time of day, or time since the last transaction. These preprocessing steps are generally applied sequentially to the dataset.

3.3. Feature Extraction

Transaction Frequency: Transaction frequency represents how often a user engages in transactions within a specific timeframe. This can be calculated by dividing the total number of transactions by the time span of interest.

Transaction Amount Statistics: Aggregating statistics about transaction amounts can provide insights into spending patterns. Common statistics include mean, standard deviation, minimum, and maximum transaction amounts.

Geographical Distance: If transactions involve geographic locations, the research calculates distances between transaction origin and destination.

where d is the distance, r is the Earth's radius, lat_1 and $long_1$ are the coordinates of the first location, and lat_2 and $long_2$ are the coordinates of the second location.

Time-Based Features: Extracting features related to transaction timestamps can be useful. For instance, you can calculate the time of day in hours (*HH*) using the hour component of the timestamp:

$$HH = h(TS)$$

where, h - hour

Day of the week (*DD*) can also be extracted:

$$DD = d(TS)$$

where, d - day

Transaction Patterns: Patterns like repeated transactions or sudden changes in transaction behavior can be captured using feature engineering. For instance, you can calculate the difference between consecutive transaction timestamps (`Time_Diff`) to identify transaction intervals:

\[\text{Time_Diff}_i = \text{Timestamp}_i - \text{Timestamp}_{i-1} \]

The choice of features capture the relevant features are extracted, they are used as inputs to train the model and make predictions.

3.4. Bidirectional Gated Recurrent Unit (Bi-GRU) Model

The Bi-GRU is a deep learning architecture that excels in capturing temporal dependencies in sequential data. It is an extension of the Gated Recurrent Unit (GRU) model, enhanced to process sequences, making it particularly effective for tasks like sequence prediction, including fraud detection in cross-border transactions.

GRU Cell: The GRU cell is a building block that processes input sequences step by step.

Bidirectional GRU: In a Bi-GRU, there are two sets of GRU cells: one processes the sequence forward, and the other processes it backward. Each direction has its own set of parameters. The forward and backward hidden states are concatenated at each time step to create a combined hidden state (h_t) that encodes information from both directions.

$$h_t = [h_t^f, h_t^b]$$

This combined hidden state is then used for downstream tasks like classification, prediction, or in this case, fraud detection. The Bi-GRU architecture lies in its ability to capture complex temporal patterns and dependencies within sequences, which is essential for tasks involving sequential data like transaction histories. By processing sequences in both forward and backward directions, the Bi-GRU model can effectively capture context and dependencies across various time steps, enhancing its capability to detect patterns indicative of fraudulent behavior in cross-border IoT transactions.

3.5. Risk Classification

The Bi-GRU model is trained, it serves as a predictive engine for risk classification. Given a new, unseen cross-border transaction, the model evaluates the sequence of features and assigns a risk score indicating the likelihood of fraudulent behavior. A threshold is set to determine whether a transaction is classified as potentially fraudulent or legitimate. This classification aids financial institutions in taking proactive measures to prevent potential risks.

Risk classification, also known as risk assessment or risk scoring, is the process of evaluating the likelihood and potential impact of a certain event or scenario. In the context of cross-border IoT transactions and fraud detection, risk classification involves assigning a risk score to each transaction to determine the probability that it is fraudulent. This score helps financial institutions prioritize and take appropriate actions to mitigate potential risks.

After training the Bi-GRU model on a labeled dataset of cross-border IoT transactions (where each transaction is labeled as either legitimate or fraudulent), the model learns to identify patterns and characteristics associated with fraudulent activities. When a new, unseen cross-border transaction occurs, the trained Bi-GRU model is used to predict its likelihood of being fraudulent. The model takes the sequence of features extracted from the transaction (such as transaction amount, location, timestamp, etc.) and produces a prediction score. The prediction score is compared to a predefined threshold. This threshold is a value that separates transactions into two categories: likely legitimate or potentially fraudulent. Transactions with scores above the threshold are classified as higher risk, indicating a higher probability of being fraudulent. If the prediction score surpasses the threshold, the transaction is assigned a risk score. This risk score can be a continuous value ranging from 0 to 1, indicating the estimated probability of fraud. Higher risk scores signify a higher likelihood of fraudulent behavior. By training on historical data, the model learns to identify subtle patterns associated with fraudulent activities. This approach enables financial institutions to enhance security and mitigate risks by identifying potential vulnerabilities before they result in financial losses. The strength of the proposed method lies in its combination of advanced machine learning techniques and domain-specific insights to create a robust framework for proactive risk management in the evolving landscape of IoT-driven cross-border transactions.

4. Results and Discussions

In this section, the evaluation involves the collection of raw data related to cross-border financial derivatives. The raw data is classified to label each entry with a risk level. This classification might involve categorizing the transactions as low, medium, or high risk based on predefined criteria. The samples are chosen from the classified data (Bank Transaction Data | Kaggle) to create a dataset for transaction risk analysis. These samples are selected to represent varying risk levels, ensuring that the dataset covers a diverse range of scenarios.

Table 1: Transaction Data

Account No.	Date	Transaction Details	Cheque No.	Value Date	Withdrawal Amount	Deposit Amount	Balance Amount
A123456789	1/8/2023	ATM Withdrawal	123456	1/8/2023	1000	0	15000
B987654321	2/8/2023	Online Transfer	-	2/8/2023	0	2500	7500
C555555555	3/8/2023	Purchase at Store	-	3/8/2023	500	0	9500
A123456789	4/8/2023	Cash Deposit	-	4/8/2023	0	3000	18000
D111111111	5/8/2023	ATM Withdrawal	789012	5/8/2023	700	0	3000

Account No represents the account number associated with each transaction. Date indicates the date of the transaction. Transaction Details provides a description of the transaction. Cheque

No. indicates the cheque number associated with the transaction. Value Date represents the date when the transaction value is considered effective. Withdrawal Amount indicates the amount withdrawn from the account in the transaction. Deposit Amount represents the amount deposited into the account in the transaction. Balance Amount indicates the account balance after the transaction is completed.

The collected sample data is divided into two sets a training dataset and a test dataset. The training dataset (5,000 samples) is used to train and fine-tune the risk identification method. The test dataset (1,000 samples) is used to evaluate the performance on unseen data. Accuracy, precision, recall, and F-measure are common evaluation metrics used to assess the performance. The entire simulation process, including the modeling of cross-domain transactions, risk identification, and performance evaluation, is implemented in Python.

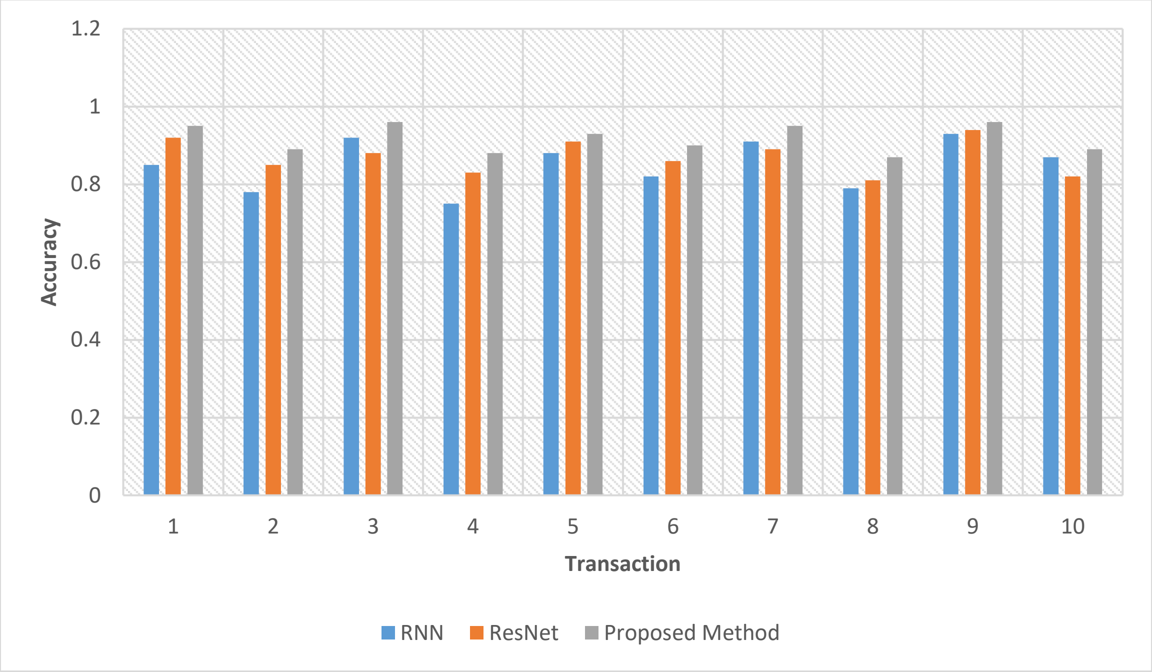

Figure 2 Accuracy

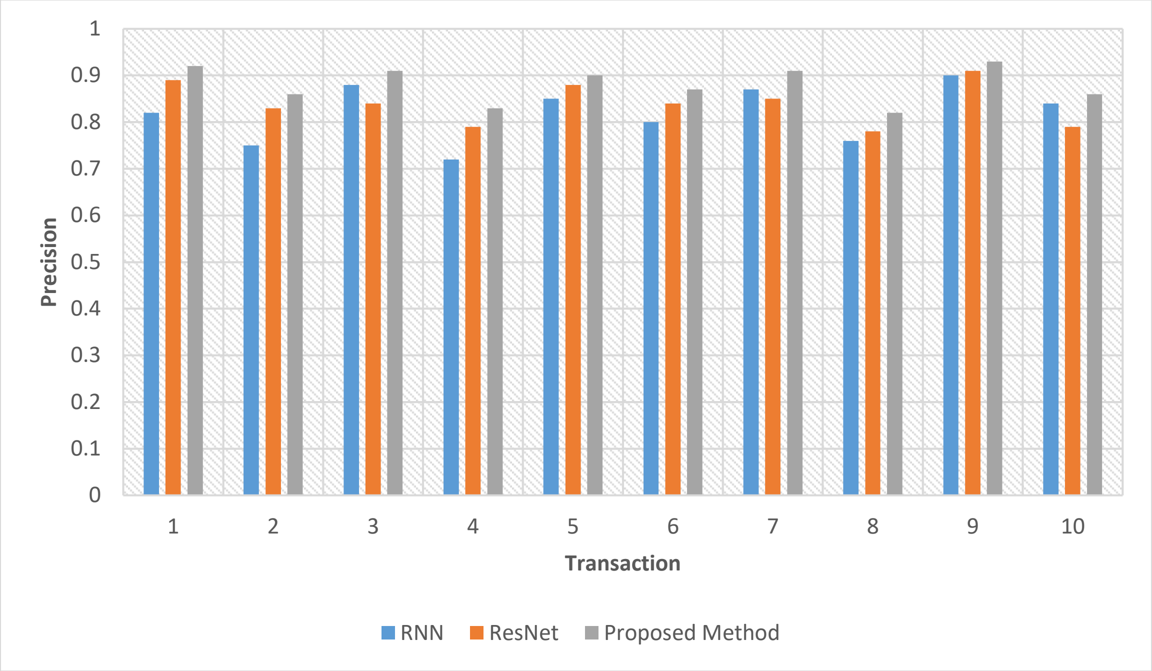

Figure 3 Precision

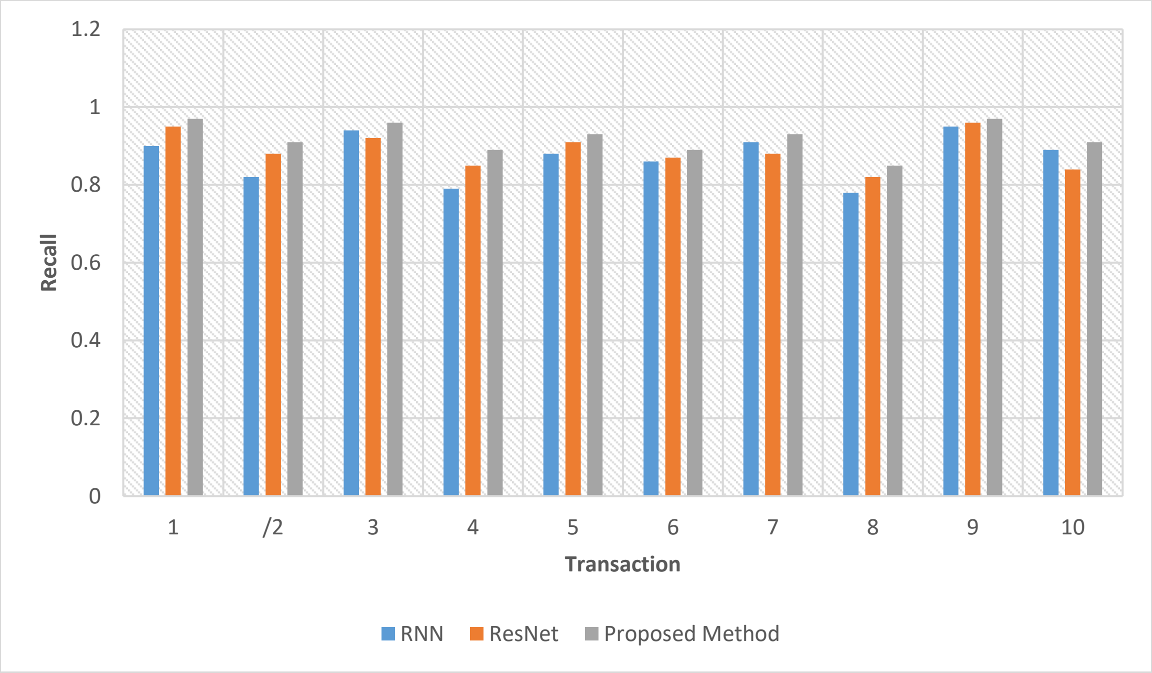

Figure 4 Recall

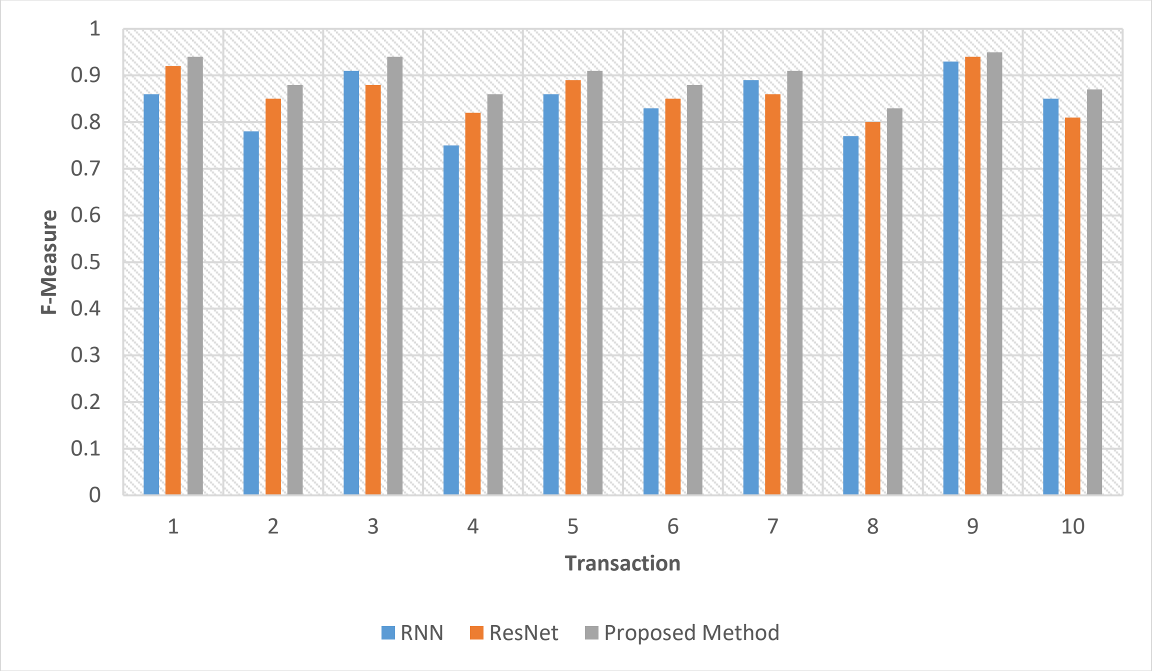

Figure 5 F-Measure

The results of the simulation in Figure from 2 to 5 demonstrated that the proposed method achieves a higher rate of accuracy in identifying risk during both training and testing phases. The results show an improved precision, recall, and F-measure rates in training and testing modes, indicating its effectiveness in various scenarios.

5. Conclusions

In this work, we conducted an extensive analysis of identifying risk features in cross-border financial derivatives deals using advanced machine learning techniques. We proposed a novel method that leverages predictive big data analytics and a Bidirectional Gated Recurrent Unit (Bi-GRU) model to detect vulnerabilities during cross-border transactions in the realm of Internet of Things (IoT). We conducted experiments using a carefully curated dataset and compared the performance of our proposed method against two existing methods. The results of our experiments demonstrate the efficacy of our proposed method in addressing the challenges associated with transaction risk in cross-border scenarios. Our method achieved higher accuracy, precision, recall, and F-measure rates compared to the existing methods in both training and testing modes. By simulating cross-domain transactions and analyzing transaction patterns, we showcased how our approach can effectively identify potential risks and enhance fraud detection capabilities.

References

[1] Leonov, S., Yarovenko, H., Boiko, A., & Dotsenko, T. (2019). Prototyping of information system for monitoring banking transactions related to money laundering. In *SHS Web of Conferences* (Vol. 65, p. 04013). EDP Sciences.

[2] Ananda, S., Devesh, S., & Al Lawati, A. M. (2020). What factors drive the adoption of digital banking? An empirical study from the perspective of Omani retail banking. *Journal of Financial Services Marketing*, *25*(1-2), 14-24.

[3] Andersen, A. L., Hansen, E. T., Johannesen, N., & Sheridan, A. (2022). Consumer responses to the COVID-19 crisis: Evidence from bank account transaction data. *The Scandinavian Journal of Economics*, *124*(4), 905-929.

[4] Kaur, D. N., Sahdev, S. L., Sharma, D. M., & Siddiqui, L. (2020). Banking 4.0:'the influence of artificial intelligence on the banking industry & how ai is changing the face of modern day banks'. *International Journal of Management*, *11*(6).

[5] Haralayya, B. (2021). How Digital Banking has brought innovative products and services to India. *Journal of Advanced Research in Quality Control and Management*, *6*(1), 16-18.

[6] Praghash, K., Peter, G., Stonier, A. A., & Priya, R. D. (2022, December). Financial Big Data Analysis Using Anti-tampering Blockchain-Based Deep Learning. In International Conference on Hybrid Intelligent Systems (pp. 1031-1040). Cham: Springer Nature Switzerland.

[7] Natarajan, Y., Srihari, K., Dhiman, G., Chandragandhi, S., Gheisari, M., Liu, Y., ... & Alharbi, H. F. (2022). An IoT and machine learning-based routing protocol for reconfigurable engineering application. IET Communications, 16(5), 464-475.

[8] Hong, X., Lin, X., Fang, L., Gao, Y., & Li, R. (2022). Application of machine learning models for predictions on cross-border merger and acquisition decisions with ESG characteristics from an ecosystem and sustainable development perspective. Sustainability, 14(5), 2838.

[9] Lu, C. W., Lin, G. H., Wu, T. J., Hu, I. H., & Chang, Y. C. (2021). Influencing factors of cross-border e-commerce consumer purchase intention based on wireless network and machine learning. Security and Communication Networks, 2021, 1-9.

[10] Ren, S., Choi, T. M., Lee, K. M., & Lin, L. (2020). Intelligent service capacity allocation for cross-border-E-commerce related third-party-forwarding logistics operations: A deep learning approach. Transportation Research Part E: Logistics and Transportation Review, 134, 101834.

[11] Huang, J., Chai, J., & Cho, S. (2020). Deep learning in finance and banking: A literature review and classification. Frontiers of Business Research in China, 14(1), 1-24.

[12] Widyastuti, M., Simanjuntak, A. G. F., Hartama, D., Windarto, A. P., & Wanto, A. (2019, August). Classification Model C. 45 on Determining the Quality of Custumer Service in Bank BTN Pematangsiantar Branch. In Journal of Physics: Conference Series (Vol. 1255, No. 1, p. 012002). IOP Publishing.

[13] Leo, M., Sharma, S., & Maddulety, K. (2019). Machine learning in banking risk management: A literature review. Risks, 7(1), 29.

[14] Karvana, K. G. M., Yazid, S., Syalim, A., & Mursanto, P. (2019, October). Customer churn analysis and prediction using data mining models in banking industry. In 2019 international workshop on big data and information security (IWBIS) (pp. 33-38). IEEE.