International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Ml-Driven Fraud Detection System to Classify Transactions as Fraudulent or Legitimate in Real-Time

Chilakalapudi Lakshmi Priya, [1]Upputhala Uday Sai[2], Buddha Tanusri[3],Mr.P.Mahesh[4]
[1,2,3] UG Scholar, Dept. ofCSE(AI&ML), St. Martin's Engineering College, Secunderabad, Telangana, India, 500100
[4]Assistant Professor, Dept. of CSE(AI&ML), St. Martin's Engineering College, Secunderabad, Telangana, India, 500100
lakshmikarthik27@gmail.com

Abstract:

The detection of suspicious financial transactions has been a critical focus in the financial industry for decades. Traditionally, financial institutions employed rule-based systems for identifying potentially fraudulent activities. These systems rely on predefined thresholds and patterns, such as large transactions or frequent deposits, to flag suspicious activities. While effective to some extent, traditional systems face significant limitations. They often generate a high rate of false positives, requiring manual intervention to review flagged transactions. Additionally, these systems struggle to adapt to evolving fraud patterns, making them less effective in detecting sophisticated financial crimes. The growing complexity and volume of financial transactions in the digital era have heightened the need for advanced detection mechanisms. Traditional systems fail to address the dynamic nature of financial fraud, leading to inefficiencies in preventing financial losses. This creates a pressing need for a more adaptable, accurate, and scalable approach to detecting suspicious transactions. The lack of adaptability in traditional methods, combined with the significant financial and reputational risks posed by undetected fraud, underscores the necessity of a more robust detection framework. The goal is to enhance the ability to detect anomalous patterns in financial data with minimal false positives while maintaining efficiency and scalability. The proposed system introduces an innovative solution that leverages an autoencoder-based model combined with a risk-based assessment strategy. This approach aims to capture subtle anomalies in transaction data that deviate from normal patterns, enabling the identification of suspicious activities. The integration of a risk-based framework ensures that the model considers contextual factors,

Keywords: Suspicious financial transactions, Fraud Detection, Manual intervention, Evolving Fraud patterns, Dynamic fraud detection, High-risk transactions, Financial fraud prevention.

1.INTRODUCTION

Suspicious financial transaction detection is a key area of concern in the financial industry, particularly in combating fraud and money laundering. The concept emerged in the 20th century with theintroduction of the first fraud detection systems, relying on manual checks and basic software tools. In India, the Reserve Bank of India (RBI) and financial institutions have been actively working towards improving fraud detection techniques. According to the Financial Intelligence Unit of India (FIU-IND), there was a significant increase in suspicious transaction reports in recent years. In 2020

alone, over 3.5 million suspicious transaction reports were filed, reflecting the growing challenge of financial fraud in the country. Traditional systems failed to keep pace with the complexity of financial crimes, including cyber fraud, phishing, and money laundering. In response, India has adopted various technological advancements to improve detection, including machine learning algorithms. The need for a more automated, accurate, and scalable solution has grown in the face of digital banking, e-commerce, and the increasing number of financial transactions. As the economy becomes more digitized, the financial sector must adapt to these changes by implementing more sophisticated detection models.

The need for an advanced suspicious financial transaction detection system is more pressing than ever in real-time financial operations. With the growth of digital banking, e-commerce, and mobile payments, the volume of financial transactions has skyrocketed, making manual fraud detection methods increasingly ineffective. Financial institutions face the challenge of processing large amounts of data quickly while identifying fraudulent activities in real time. Traditional systems often fail to keep up, leading to delays in fraud detection and financial losses. Additionally, the rise of advanced fraud techniques such as synthetic identity fraud and money laundering demands more sophisticated detection methods. Machine learning models can address these challenges by providing real-time, adaptive, and scalable solutions that detect anomalies and fraud with higher accuracy.

The real-time nature of this project allows for immediate action to be taken against suspicious transactions, reducing the potential for financial damage and enhancing the trustworthiness of financial systems. In India, where the financial sector is rapidly expanding, there is an urgent need for such systems to protect consumers and institutions alike.

2. LITERATURE SURVEY

Singh & Best proposed a method using data visualization to detect suspicious activities for anti-money laundering (AML). They demonstrated how data visualization techniques could effectively identify unusual transaction patterns indicative of money laundering, offering a powerful tool for monitoring financial systems and improving compliance procedures in the banking sector. Whisker & Lokanan examined the risks posed by mobile money in the context of anti-money laundering and counter-terrorist financing. They discussed the challenges that mobile money presents, including its anonymous nature and the ease with which it can be exploited for illicit financial activities, and recommended strategies to mitigate these risks. Dobrowolski & Sułkowski presented a sustainable model for implementing anti-money laundering measures aligned with the United Nations' development goals. They proposed an integrated approach that emphasizes sustainable financial systems and the role of AML in promoting global economic stability and development, while addressing the challenges of compliance in varying regulatory environments. Irwin et al. analyzed typologies of money laundering and terrorism financing, focusing on the techniques and methods used to conceal illicit financial activities.

They explored patterns and characteristics of these criminal activities, providing a comprehensive overview of how

such practices can be detected and prevented through effective AML policies. Uthayakumar et al. introduced a framework based on swarm intelligence for classification rule induction, applied to bankruptcy prediction and credit risk analysis. They showed how the application of machine learning techniques like swarm intelligence could improve prediction accuracy in financial sectors, particularly for detecting risks associated with money laundering. KOFIU presented the risk-based approach (RBA) standards for anti-money laundering (AML) and counter-terrorism financing (CFT) in financial investment businesses. This work highlighted how financial institutions can apply RBA to identify high-risk activities and tailor their compliance efforts accordingly, thereby optimizing the use of resources in combating financial crimes

They discussed the integration of various regulatory frameworks and the challenges involved in creating a comprehensive system that effectively prevents money laundering while ensuring financial system stability. Pavlidis addressed the unintended consequences of implementing anti-money laundering standards, specifically those set by the Financial Action Task Force (FATF). The research examined how the FATF's strict compliance requirements may sometimes have negative effects, such as increasing financial exclusion, and suggested ways to mitigate these impacts. Celik explored the impact of the FATF's recommendations on financial inclusion, drawing insights from mutual evaluations and national risk assessments. The study emphasized the balancing act between enforcing AML/CFT regulations and maintaining access to financial services, particularly for underserved communities.

Jayasekara discussed the challenges of implementing an effective risk-based supervision approach for AML and countering the financing of terrorism under the 2013 FATF methodology. The paper focused on the difficulties financial institutions face when adopting these complex systems and provided recommendations to enhance supervisory effectiveness. Raghavan explored the integration of anti-money laundering practices into corporate governance and finance functions, with a focus on compliance with the Bank Secrecy Act (BSA) and anti-money laundering (AML) regulations. The study showed how AML requirements are reshaping corporate governance, emphasizing the need for stricter internal controls to prevent financial crimes. Raghavan examined how AML regulations are influencing corporate governance structures and finance functions in financial institutions.

He discussed the integration of AML practices into organizational frameworks and highlighted the evolving nature of compliance requirements, particularly in response to emerging financial crimes. Labib et al. surveyed machine learning approaches for anti-money laundering (AML) and counter-terrorism financing techniques. They reviewed various ML techniques, discussing how they can be applied to detect suspicious transactions, identify potential money laundering activities, and improve overall AML effectiveness. Cherif et al. conducted a systematic review of credit card fraud detection methods using disruptive technologies. They focused on how advancements in AI and machine learning

have significantly enhanced the accuracy of fraud detection systems, especially in the context of financial transactions, which are often targeted by money laundering activities.

Senator et al. proposed an AI-based system designed to identify potential money laundering activities from large cash transaction reports. Their research introduced the Financial Crimes Enforcement Network's (FinCEN) AI system for detecting suspicious patterns, emphasizing the role of AI in automating the detection of financial crimes. Wang & Yang introduced a decision tree-based method for evaluating money laundering risks. Their model used historical transaction data to classify transactions based on their risk levels, enabling more effective detection and prevention of money laundering activities. Zhang & Zhou explored the use of data mining techniques in financial applications, focusing on how these methods can uncover hidden patterns in transaction data that may indicate illicit activities such as money laundering. They discussed the potential of data mining to enhance financial monitoring systems and improve the identification of financial crimes.

3. PROPOSED METHODOLOGY

An **Autoencoder** is an unsupervised learning algorithm used primarily for dimensionality reduction or feature learning. It consists of two parts: an **encoder**, which compresses the input data into a latent-space representation, and a **decoder**, which reconstructs the input data from this representation. The purpose of an autoencoder is to learn an efficient representation of the input data in a lower-dimensional space.

When combined with a **Random Forest Classifier (RFC)**, the autoencoder is used to reduce the dimensionality of the data by learning a compressed representation of the features. The output of the autoencoder's encoder is then used as input to train an RFC for classification tasks. This combination is effective in detecting fraud because the autoencoder can highlight anomalies in the data, while the RFC helps to classify whether atransaction is fraudulent or not based on thosefeatures.

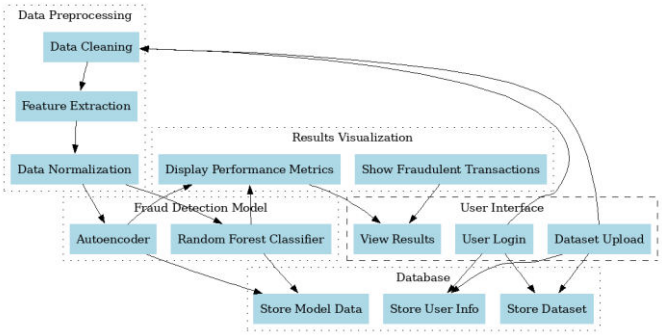


Fig.3.1: Architectural Block Diagram of The Proposed System.

Architecture of Autoencoder with RFC

- **Encoder**: The encoder takes the raw data (e.g., transaction details) and transforms it into a compact, lower-dimensional representation.
- **Latent Space**: The compressed representation of the data learned by the encoder is the latent space.
- **Decoder**: The decoder reconstructs the input data from the latent space. This step ensures that the autoencoder has learned an efficient representation.
- **Random Forest Classifier (RFC)**: The reduced features from the autoencoder's encoder are used as input for the RFC. The RFC consists of multiple decision trees that collectively vote on the class label, whether a transaction is fraudulent or not.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Fraud detection in financial transactions is a critical task that requires a robust and systematic approach to identify anomalies and prevent financial losses. This research outlines a step-by-step methodology for developing and comparing two machine learning models: an existing Deep Neural Network (DNN) model and a proposed Autoencoder-Random Forest (AERF) model. The procedure involves dataset preparation, preprocessing, model training, and performance evaluation. Below are the detailed steps of this research.

1. **Autoencoder for Feature Extraction:**
   - The Autoencoder is an unsupervised neural network designed to reconstruct input data.
   - Architecture: It consists of an encoder (compresses input into a lower-dimensional representation) and a decoder (reconstructs the input from the compressed features).
   - Training: The Autoencoder is trained to minimize reconstruction loss, measured by mean squared error (MSE). Once trained, the encoder part is extracted and used to transform the original features into a compact, informative representation.
2. **Random Forest Classifier:**
   - The transformed features from the Autoencoder are used as input for the Random Forest Classifier.
   - Random Forest is an ensemble learning method that combines multiple decision trees to improve classification performance and reduce overfitting.
   - Hyperparameter Tuning: Parameters like the number of trees and maximum depth are optimized to achieve the best performance.

4. EXPERIMENTAL ANALYSIS

The homepage serves as the initial interface for users and administrators. It provides navigation options for various functionalities like registration, login, and dataset upload. The design ensures an intuitive and user-friendly experience. The figure 2 represents the registration page, which is a shared form for both users and administrators. The page captures essential information such as username, password, and other identification details to create a new account in the system.

Fig 1: Home Page of the Financial Transaction Detection.

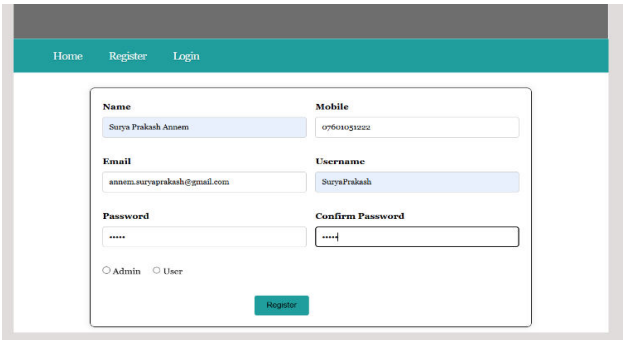


Fig 2: Common Registration for user and admin.

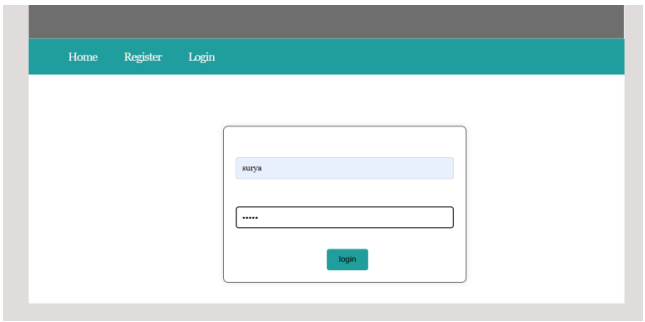
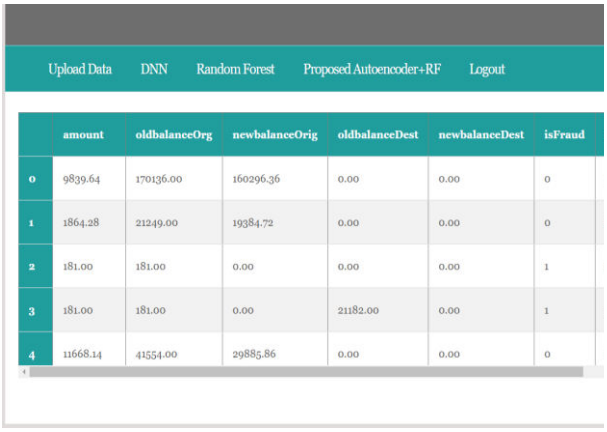


Fig 3: User login for using Transaction detection.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

The user login page allows registered users to access the fraud detection system. It includes fields for entering credentials like username and password. Upon successful login, users can utilize the system's features, such as uploading datasets for fraud detection.



	amount	oldbalanceOrg	newbalanceOrig	oldbalanceDest	newbalanceDest	isFraud
0	9839.64	170136.00	160296.36	0.00	0.00	0
1	1864.28	21249.00	19384.72	0.00	0.00	0
2	181.00	181.00	0.00	0.00	0.00	1
3	181.00	181.00	0.00	21182.00	0.00	1
4	11668.14	41554.00	29885.86	0.00	0.00	0

Fig 4: Sample Fraud Transaction Uploaded Dataset.

This figure depicts a sample of the dataset uploaded by users or administrators. It contains financial transaction details, including attributes such as transaction step, type, source and destination balances, fraud indicators, etc., to analyze and detect fraudulent activities.

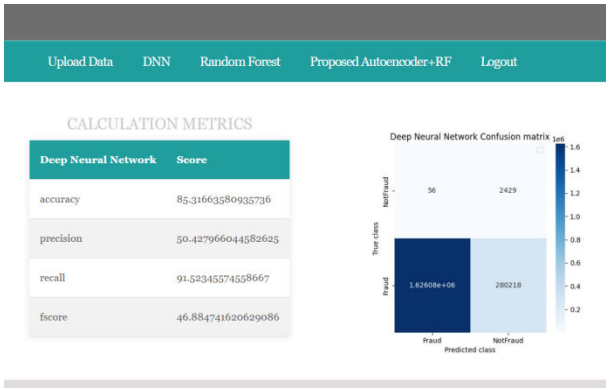


Fig 5: Performance metrics of the Existing DNN model.

This figure highlights the performance metrics of the Deep Neural Network (DNN) model applied to the financial transaction dataset. Metrics include:

- **Accuracy**: 85.31%
- **Precision**: 50.43%
- **Recall**: 91.52%
- **F1-Score**: 46.88%

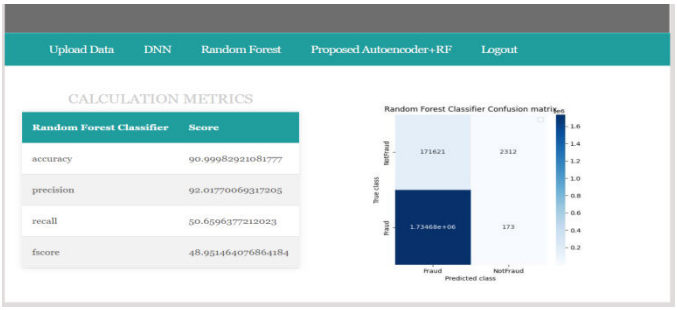


Fig 6: Performance metrics of the Existing RFC model.

This figure shows the performance metrics of the Random Forest Classifier (RFC) model. Metrics include:

- **Accuracy**: 90.99%
- **Precision**: 92.02%
- **Recall**: 50.66%
- **F1-Score**:48.95%

The RFC model outperforms DNN in accuracy and precision, making it more reliable for correctly identifying fraudulent transactions.

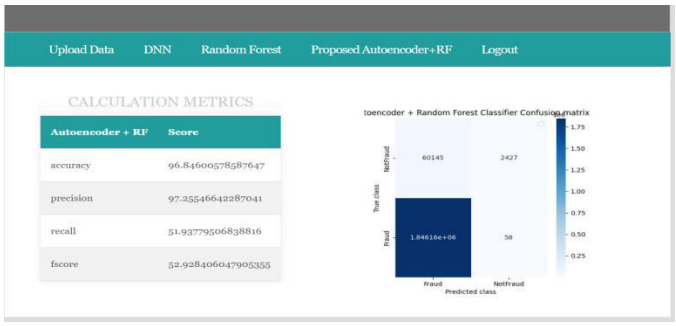


Fig 7: Performance metrics of the Proposed Auto Encoder + RFC model.

This figure presents the metrics for the hybrid Autoencoder + Random Forest Classifier model. Metrics include:

- **Accuracy**: 96.85%
- **Precision**: 97.26%
- **Recall**: 51.94%
- **F1-Score**:52.93%

The proposed model achieves the highest accuracy and precision, demonstrating significant improvements in detecting fraudulent transactions while maintaining a balanced F1-score.

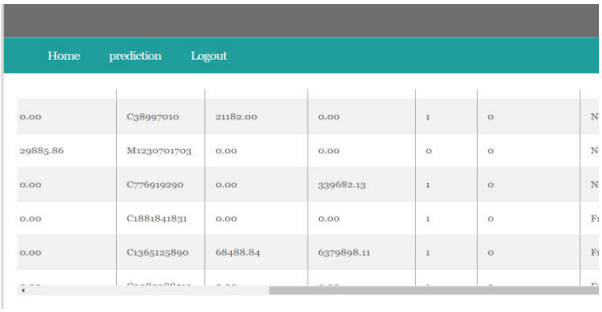


Fig 8: Proposed model prediction on user uploaded test data.

This figure illustrates the proposed model's predictions on a sample user-uploaded test dataset. The Autoencoder + RFC model effectively identifies fraudulent transactions, showcasing its applicability to real-world datasets.

5. CONCLUSION

The Research combines the strengths of Autoencoders for feature extraction and dimensionality reduction with the Random Forest Classifier (RFC) for robust classification to detect fraudulent transactions in financial datasets. This hybrid approach uses Autoencoders to detect hidden patterns in transaction data and RFCs for accurate and interpretable classification. The system's architecture ensures scalability, efficiency, and accuracy in fraud detection, addressing challenges posed by imbalanced datasets and complex transactional behaviours. The project significantly contributes to reducing financial losses and enhancing trust in financial systems.

1. **Integration with Real-Time Systems:** The model can be integrated into live transaction monitoring systems for real-time fraud detection, enabling proactive responses to suspicious activities.
2. **Enhancing Model Performance**: Advanced techniques like ensemble learning with gradient-boosted models or transformers could be explored to improve prediction accuracy and computational efficiency further.
3. **Explainability and Interpretability**: Implementing explainable AI (XAI) techniques to make the fraud detection decisions more transparent for stakeholders, ensuring better understanding and trust in the system.
4. **Application to Multimodal Data:** Incorporating additional data sources like user behavioral patterns, geolocation, and device information could enhance the model's ability to identify fraudulent transactions.
5. **Domain Adaptation:** Extending the system to other industries, such as insurance or healthcare, to detect fraud in claims and billing data.
6. **Adaptation to Emerging Threats:** Regular updates to the system using continuous learning methods to adapt to new fraud techniques and patterns as they evolve.

REFERENCES

[1]. Singh, K., & Best, P. (2019). Anti-money laundering: Using data visualization to identify suspicious activity. International Journal of Accounting Information Systems, 34, 100418.

[2]. Whisker, J., & Lokanan, M. E. (2019). Anti-money laundering and counter-terrorist financing threats posed by mobile money. Journal of Money Laundering Control, 22(1), 158-172.

[3]. Dobrowolski, Z., & Sułkowski, Ł. (2019). Implementing a sustainable model for anti-money laundering in the United Nations development goals. Sustainability, 12(1), 244.

[4]. Samantha Maitland Irwin, A., Raymond Choo, K. K., & Liu, L. (2011). An analysis of money laundering and terrorism financing typologies. Journal of Money Laundering Control, 15(1), 85-111.

[5]. Uthayakumar, J., Vengattaraman, & Dhavachelvan, T. P. (2022). Swarm intelligence-based classification rule induction (CRI) framework for qualitative and quantitative approach: An application of bankruptcy prediction and credit risk analysis. Journal of King Saud University - Computer and Information Sciences, 32(6), 647-657.

[6]. KOFIU. (2017). Risk-Based Approach (RBA) Processing Standards for AML/CFT in Financial Investment Businesses. Institutional Operations Division, Financial Intelligence Unit.

[7]. Lee, C.-J., & Lee, J.-C. (2013). Experiences and methodology of Korea's anti-money laundering system deployment and development. Knowledge Sharing Program: KSP Modularization.

[8]. Pavlidis, G. (2023). The dark side of anti-money laundering: Mitigating the unintended consequences of FATF standards. Journal of Economic Criminology, 100040.

[9]. Celik, K. (2021). Impact of the FATF Recommendations and their Implementation on Financial Inclusion: Insights from Mutual Evaluations and National Risk Assessments.

[10]. Jayasekara, S. D. (2018). Challenges of implementing an effective risk-based supervision on anti-money laundering and countering the financing of terrorism under the 2013 FATF methodology. Journal of Money Laundering Control, 21(4), 601-615.

[11]. Raghavan, K. R. (2006). Integrating anti-money laundering into the compliance structure: How the requirements for compliance with BSA/AML are changing the emphasis of corporate governance and finance functions. Bank Accounting & Finance, 19(6), 29-37.

[12]. Schott, P. A. (2006). Reference Guide to Anti-Money Laundering and Combating the Financing of Terrorism. The International Bank for Reconstruction and Development/The World Bank/The International Monetary Fund.

[13]. Labib, N. M., Rizka, M. A., & Shokry, A. E. M. (2020). Survey of machine learning approaches of anti-money laundering techniques to counter terrorism finance. In Internet of Things—Applications and Future: Proceedings of ITAF 2019 (pp. 73-87). Springer.

[14]. Cherif, A., Badhib, A., Ammar, H., Alshehri, S., Kalkatawi, M., Imine, A.

(2023). Credit card fraud detection in the era of disruptive technologies: A systematic review, Journal of King Saud University - Computer and Information Sciences, 35(1), 145-174.

[15]. Senator, T. E., Goldberg, H. G., Wooton, J., Cottini, M. A., Khan, A. U., Klinger, C. D., Llamas, W. M., Marrone, M. P., &Wong, R. W. (1995). Financial crimes enforcement network AI system (FAIS) identifying potential money laundering from reports of large cash transactions. AI magazine, 16(4), 21-21.

[16]. Wang, S.-N., & Yang, J.-G. (2007). A money laundering risk evaluation method based on decision tree. 2007 international conference on machine learning and cybernetics, January.

[17]. Zhang, D., & Zhou, L. (2004). Discovering golden nuggets: Data mining in financial application. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 34(4), 513-522.

[18]. Schott, P. A. (2006). Reference Guide to Anti-Money Laundering and Combating the Financing of Terrorism. The World Bank.

[19]. ICAI. (2013). Importance of Anti-Money Laundering Measures and Effective KYC in Financial Transactions. Institute of Chartered Accountants of India.

[20]. Geiger, H., & Wuensch, O. (2007). The Fight Against Money Laundering: An Economic Analysis of a Cost-Benefit Paradoxon. Journal of Money Laundering Control, 10(1), 91-105