UTS UNIVERSITY OF TECHNOLOGY SYDNEY

Explainable Deep Learning Approach for Detecting Money Laundering Transactions in Banking System

by Dattatray Vishnu Kute

Thesis submitted in fulfilment of the requirements for the degree of

Master of Science (Research) in Computing Sciences

under the supervision of
Distinguished Professor Biswajeet Pradhan and
Dr. Nagesh Shukla

University of Technology Sydney
Faculty of Engineering and Information Technology
Centre for Advanced Modelling and Geospatial Information Systems

November 2022

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, **Dattatray Vishnu Kute** declare that this thesis, is submitted in fulfilment of the requirements for the award of **Master of Science (Research) in Computing Sciences**, in the **Faculty of Engineering and Information Technology** at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Signature:

Production Note:
Signature removed prior to publication.

Dattatray Vishnu Kute

Date: 10 December 2022

ACKNOWLEDGEMENT

I would like to express my sincere gratitude and respect to my esteemed principal supervisor, Distinguished Professor Biswajeet Pradhan and my co-supervisor Dr Nagesh Shukla. Their achievements in scientific world has been a continuous inspiration for me. Their continuous guidance throughout the research duration, all the way from assessing the research questions, defining research methodology, validating the experiments, critical reviews, and suggestions helped me immensely. Each review comment backed up by so much experience was always fascinating to see how it was changing the direction or flow of the research, and scientific papers that we published during my research period. Without their constant encouragement and supervision, this research would not have been possible.

I would like to extend my sincere gratitude to the candidature assessment panellists who patiently listened to my research and provided valuable feedback to shape up the research.

I am very grateful to the Centre for Advanced Modelling and Geospatial Information Systems, Faculty of Engineering and Information Technology, the University of Technology Sydney for offering research scholarships that have enabled me to accomplish this study.

From industry, I offer my special thanks to Mr. Girish Rao, Finance Crime and AML industry expert, who provided the guidance on AML domain and also helped to validate the findings of several experiments from practical application point of view.

I thank my employer TATA Consultancy Services and my managers for encouraging and supporting me throughout the research duration.

I want to thank my peers Anirban Adak and Abhirup Dikshit who had given their time to brainstorm several aspects of my research that helped to shape and sharpen my thoughts.

I express my deep gratitude to my parents for their continuous encouragement and blessings, my wife Swati for untiring support, and my son Shouryan for sacrificing several weekends for my study.

I also thank everyone whose name is not included here but have helped me directly or indirectly.

LIST OF PAPERS/PUBLICATIONS

Following papers are produced as a part of the research:

- Kute, D. V., B. Pradhan, N. Shukla, and A. Alamri. 2021. 'Deep Learning and Explainable Artificial Intelligence Techniques Applied for Detecting Money Laundering-A Critical Review', IEEE Access, 9: 82300-17 - **Published**

- Kute, D. V., B. Pradhan, N. Shukla, and A. Alamri. 2022. 'Explainable Deep Learning Model for Predicting Money Laundering Transactions', – **Under review**

Note: Thesis includes the contents from the papers published above.

TABLE OF CONTENTS

CERTIFICATE OF ORIGINAL AUTHORSHIP ..... **i**

ACKNOWLEDGEMENT ..... **ii**

LIST OF PAPERS/PUBLICATIONS ..... **iii**

LIST OF TABLES ..... **vii**

LIST OF FIGURES ..... **viii**

ABSTRACT ..... **ix**

1 INTRODUCTION **1**

1.1 General Introduction 1
1.2 Research Background 2
1.3 Research Gaps 7
1.4 Problem Statement 8
1.5 Research Questions 8
1.6 Research Aim and Objectives 9
1.7 Scope of Thesis 9
1.8 Motivation behind Research 10
1.9 Novelty and Main Contribution 11
1.10 Thesis Organisation 12

2 LITERATURE REVIEW **14**

2.1 Introduction 14
2.2 Literature Review Methodology 16
2.2.1 Aim and Research Questions 18
2.2.2 Search and Selection Process 19
2.3 Various Techniques in Anti-Money Laundering 23
2.3.1 Machine Learning Techniques 24
2.3.2 Deep Learning Techniques 33
2.3.3 Interpretability Analysis 37
2.3.4 Explainable AI Techniques 50
2.4 Discussion and Recommendations 54
2.4.1 Strengths and Limitations of Applied Models 55
2.4.2 Current Research Challenges and Constraints 56

2.4.3 Barriers for Adoption of AI/ML in AML Domain 59
2.4.4 Future Research Directions 60
2.5 Summary 62

3 MATERIALS AND RESEARCH METHODOLOGY 64

3.1 Introduction 64
3.2 Overall Methodology 64
3.3 Data Acquisition 65
3.3.1 Synthetic Data Creation Approach 67
3.3.2 Identify Data Attributes 68
3.3.3 Define Data Creation Rules 74
3.3.4 Define Money Laundering Scenarios 75
3.3.5 Design Synthetic Data Generation Software 76
3.3.6 Generate Data Using Utility 81
3.3.7 Define Data Enrichment Rules 82
3.4 Objective 1 – Detect Suspicious Transactions 83
3.4.1 Convolutional Neural Network 83
3.4.2 Random Forest 87
3.4.3 Support Vector Machine 87
3.4.4 Extreme Gradient Boosting 88
3.4.5 Evaluation and Performance Metrics 88
3.5 Objective 2 – Explain Predictions 91
3.6 Software for Model Implementation 93
3.6.1 Modelling Software 93
3.6.2 Data Creation Software 94
3.7 Summary 94

4 RESULTS AND DISCUSSION 96

4.1 Introduction 96
4.2 Synthetic Data 96
4.3 Objective 1 - Detect Suspicious Transactions 104
4.3.1 Suspicious Transaction Detection using CNN 105
4.3.2 Validation of CNN Results using RF 109
4.3.3 Validation of CNN Results using SVM 111
4.3.4 Validation of CNN Results using XGBoost 113
4.3.5 Discussion 116

4.4 Objective 2 - Explain Predictions 120
4.4.1 Explain CNN Predictions using SHAP 120
4.4.2 Explain CNN Predictions using LIME 125
4.5 Summary 126

5 CONCLUSIONS AND FUTURE WORK RECOMMENDATIONS 128

5.1 General Conclusion 128
5.2 Conclusion of Objective 1 129
5.3 Conclusion of Objective 2 130
5.4 Practical Implications 130
5.5 Research Drawbacks and Limitations 132
5.6 Recommendations for Future Work 132

REFERENCES 135

APPENDIX A – MACHINE LEARNING HYPER PARAMETERS 150

APPENDIX B – INTERPRETATION EXAMPLES USING SHAP 152

APPENDIX C – KEY TERMINOLOGIES 153

APPENDIX D – DATA GENERATION PROCESS 155

LIST OF TABLES

Table 2.1 – Literature search queries and results 19
Table 2.2 – Literature classification 22
Table 2.3 – Machine learning methods used in AML domain 25
Table 2.4 – Review papers – machine learning in AML domain 29
Table 2.5 – Interpretability of methods used for AML domain 39
Table 2.6 – XAI review papers 51
Table 3.1 – Customer demography data attributes 70
Table 3.2 – Account data attributes 71
Table 3.3 – Transactions data attributes 72
Table 3.4 – Suspicious money laundering transactions 75
Table 3.5 – Legitimate overlapping transactions 76
Table 3.6 – Modelling software used 93
Table 3.7 – Data creation software used 94
Table 4.1 – Dataset specifications 97
Table 4.2 – Raw customer demography data sample 98
Table 4.3 – Raw accounts data sample 99
Table 4.4 – Raw transactions data sample 100
Table 4.5 – Final dataset sample 101
Table 4.6 – CNN architecture hyper parameters 106
Table 4.7 – CNN prediction results 107
Table 4.8 – CNN results comparison with RF results 109
Table 4.9 – CNN results comparison with SVM results 111
Table 4.10 – CNN results comparison with XGBoost results 114
Table 4.11 – Prediction Results CNN, RF, XGBoost and SVM 117
Table 4.12 – Original transaction records associated with CNN predictions 124

LIST OF FIGURES

Figure 1.1 – Domains in finance crime 3
Figure 1.2 – Money laundering and AML process 5
Figure 1.3 – AML ecosystem 6
Figure 2.1 – AML Process 15
Figure 2.2 – Literature review methodology 17
Figure 2.3 – Literature analysis – by article type 20
Figure 2.4 – Literature analysis – year wise articles 21
Figure 2.5 – Literature analysis – articles by journals 21
Figure 2.6 – Year wise machine learning methods published in AML domain 48
Figure 2.7 – Interpretability of machine learning methods used in AML solutions 48
Figure 2.8 – Category of machine learning techniques used for AML solutions 49
Figure 2.9 – Type of data used for model training and evaluation 49
Figure 3.1 – Research development methodology 65
Figure 3.2 – Synthetic data creation methodology 68
Figure 3.3 – Initial data attributes 69
Figure 3.4 – Synthetic data generation flow 77
Figure 3.5 – Synthetic data generation logic 79
Figure 3.6 – CNN architecture 85
Figure 4.1 – ROC curve for CNN Model 108
Figure 4.2 – ROC curve – (a) CNN and (b) RF 110
Figure 4.3 – Global interpretation of RF classifier 111
Figure 4.4 – RoC curve – (a) CNN and (b) SVM 112
Figure 4.5 – Global interpretation of SVM classifier 113
Figure 4.6 – ROC curve – (a) CNN and (b) XGBoost 115
Figure 4.7 – Global interpretation of XGBoost classifier 115
Figure 4.8 – ROC curve – (a) CNN, (b) RF, (c) XGBoost and (d) SVM 118
Figure 4.9 – SHAP force plot for suspicious transaction 123
Figure 4.10 – SHAP force plot for legitimate transaction 123

ABSTRACT

Money laundering has been a global issue since last few decades. Intergovernmental organisation, government, and reporting entities involved in cash transactions are combating against the finance crime and money laundering. The continuous change in the technology, fraud patterns and regulatory landscape is making it even harder. To automate time-consuming operations and assist the investigation officers, Artificial Intelligence (AI) based decision support systems are being developed. These systems suffer from a high false negative rate of alerts raised by the traditional machine learning methods used in the Anti-Money Laundering (AML) domain. Further non-explainability of the predictions made by black-box machine learning models and inadequate trust in the AI system slows down the adoption of AI systems in the banking sector.

This thesis presents a study on the application of Convolutional Neural Network (CNN) method for predicting suspicious money laundering transactions and explaining the CNN predictions using SHapley Additive exPlanations (SHAP) eXplainable AI (XAI) method. The CNN model is trained and tested on a synthetic data produced by authors, and compared the results with Random Forest (RF), Extreme Gradient Boosting (XGBoost) and Support Vector Machine (SVM) methods.

The results showed that CNN model outperformed other models in detecting large number of suspicious transactions, which is a good indication from compliance risk reduction point of view. On the contrary, CNN model showed higher number of false positives compared to other models which indicates more operational efforts. From the Banks perspective, reducing the compliance risk is more important than operational efforts, hence more weightage was given to Recall metric than Precision to calculate the $f_\beta$score with $\beta = 3$ to measure the performance. The scores were CNN - 78.23%, XGB – 62.09%, RF – 61.09%, and SVM - 30.86%. The SHAP XAI method applied on CNN model to interpret the predictions made by CNN model. The SHAP force plot could show relevant indicators such as cash transaction, risk rating, ‘Know Your Customer’ (KYC) state, target country and age, that explains the reasoning behind the predictions made by CNN at an individual prediction level. The key finding from the research is, the CNN model is able to detect injected suspicious money laundering transactions (or less false

negatives) very well as compared to other traditional machine learning models, and XAI SHAP method is able to bring out the interpretation of decisions made by CNN model.

Keywords: Money laundering, anti-money laundering, finance crime, deep learning, convolutional neural network, CNN, explainable artificial intelligence, XAI, SHAP

1 INTRODUCTION

This chapter provides the general introduction and background to the money laundering domain, describes the problems, research gaps, and defines the scope of this thesis by establishing the research questions, aims and objectives. Further chapter 1 describes the motivation behind conducting this research and the main contribution to scientific knowledge repository. Finally the chapter describes the thesis organisation structure.

1.1 General Introduction

Money laundering and terrorism financing has been one of the major threat to the integrity of international financial systems since last five decades (Force 2020). United Nations Office on Drugs and Crime (UNODC) estimates that the amount of money laundered every year is approximately 2 – 5% of global GDP which amounts to $800bn - $2 trillion (Crime 2021) and it is one of the biggest threat for global economy and its security (Milind Tiwari 2020). To fight against finance crime, many countries have laid the stringent Anti-Money Laundering / Counter-Terrorism Financing (AML/CTF) policies for organisations (or reporting entities) who deal in the cash, bullion, cryptocurrencies and financial transactions. For example, 'AML/CTF Act 2006' in Australia (Office of Parliamentary Counsel 2019), 'Bank Secrecy Act 1970' in USA (Network 2021). The regulatory body (such as AUSTRAC for Australia) requires the reporting entities (such as Banks, Foreign Exchange Services etc.) to deploy effective AML controls to monitor and report cash transactions beyond threshold values, International Fund Transfer Instructions (IFTIs), suspicious transactions of any kind and many more (Crime 2021). This thesis is presented by considering the reporting entity as Banks, hence all discussions, use cases will be around the banking domain.

There have been several cases where financial institutions are penalised for not maintaining effective controls to fight against the money laundering. In 2009, Credit Suisse Group of Switzerland was fined $536 million (Department of Justice 2009), in the same year Lloyds Banking Group of UK was fined $350 million (James Quinn 2009). In 2012, HSBC was fined by US authorities a sum of $1.9bn in a settlement over money

laundering (News 2012), in the same year ING Bank group of Netherlands was fined a penalty of $619 million for enabling launders to illegally move billions of dollars through US banking system (Journal 2012). In 2017, the Commonwealth Bank of Australia was ordered to pay the penalty of $700 million under AML/CTF act (Centre 2018). In 2019, 58 AML penalties handed out globally totalling $8.14bn, which is double the amount of penalty that was handed out in 2018 (Monroe 2020). Very recently, in 2020, the Westpac Bank of Australia was ordered to pay a record-breaking $1.3 billion fine for breaching countries AML/CTF law (Centre 2020). The penalties handed over to various financial institutes until today, indicates that the existing systems and controls are not adequate and effective to fight against the finance crime.

To combat the finance crime and comply with AML requirements, traditionally Banks have been using rule-based AML systems (Chen et al. 2018c), which helps to identify cash transactions beyond certain threshold values, international fund transfers and suspicious transactions. These systems detect the suspicious money laundering transactions based on the predefined rules and raise the alerts. Compliance officer investigates these alerts and if it is found to be positive, a Suspicious Matter Report (SMR) is prepared and submitted to regulatory to ensure compliance reporting. Banks have started developing AI systems, but it results in high false positive rate primarily due to usage of shallow learning methods that plateau at a certain level in spite of growth in the data. The false positive rate of these alerts is estimated over 98 % (Company 2019), which results into a huge operational cost to banks, and missing out on detecting the true money laundering transactions creates a huge risk for organisation from a regulatory compliance point of view. If the compliance risk is not managed well, it leads to heavy penalties for the financial institutions.

1.2 Research Background

In a financial institutions, finance crime includes different types of fraud as shown in **Figure 1.1**. The key focus and scope of this thesis is money laundering type of fraud.

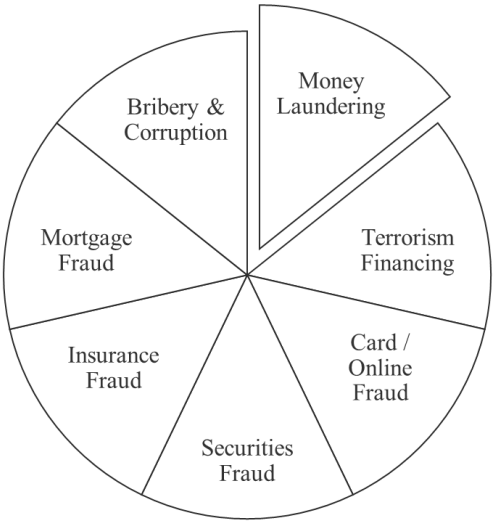


Figure 1.1 – Domains in finance crime

Money laundering is a criminal activity used to disguise the illegally obtained money and make them appear legitimate in the system. As shown in **Figure 1.2**, typically money laundering involves 3 steps (Force 2021), first step is "*placement*" where cash is introduced in the financial system, second step is "*layering*" where the complex financial transactions are performed to disguise the illegal source of the cash and third step is "*integration*" where the benefits are acquired from the transactions of illegal funds. The best step to identify a suspicious transaction is a placement step, because beyond the first step it becomes complex to trace the transactions since the transactions are layered within and/or outside of the bank purview.

AML is a set of activities performed by reporting entities to monitor the transactions and report the suspicious activities to regulatory. From AML perspective, banks use the rule-based system for transaction monitoring that raises the alerts for suspicious transaction activities (Finance 2018). These alerts are investigated by AML officers to find out enough evidences to qualify the alert as true suspicious activity, using which a Suspicious Matter Report (SMR) is prepared and submitted to Financial Intelligence Units (FIU) (Finance 2018). FIU investigates the SMR in a bigger context having access to many other data sources such as tax office, police departments, and financial intelligence data. If the FIU investigation finds adequate evidences for the SMR being suspicious, a criminal case is prepared and it is handed over to Law & Enforcement agencies for further

investigation and finding the truth. Banks are obliged to monitor the transactions and report the suspicious activities to regulatory to ensure compliance with AML/CTF policy. Effective AML systems, controls, practices are critical to manage the compliance risk.

Figure 1.3 presents the ecosystem of AML that includes customers who performs regular banking activities and complying with banking policies such as updating the identification details.

The Banks are involved in providing the financial services to customer and to comply with regulatory they are expected to monitor the transactions to identify suspicious activities and report to regulatory. The regulatory or FIU who investigates the suspicious activities reported by Banks and further involve law and enforcement agencies as required for further investigation of case. If the law and enforcement agencies are involved then a criminal investigation matter goes to court for final decision.

One of the key activities in ant-money laundering is to detect the suspicious activity. The suspicious money laundering transaction detection is a multi-step classification problem, because the confidence level of transaction being suspicious is built upon the investigation performed at each step from Alert to SMR and beyond the purview of financial institution. The decision of transaction being suspicious goes through 5 steps - alert generation system (mostly automation), manual investigation at financial institution, investigation at FIUs, investigation by Law and Enforcement agencies and finally at the court. Considering the efforts involved by various parties (Bank, FIU and Law & Enforcement agencies) it is important to get the detection of suspicious transaction correct to reduce the efforts.

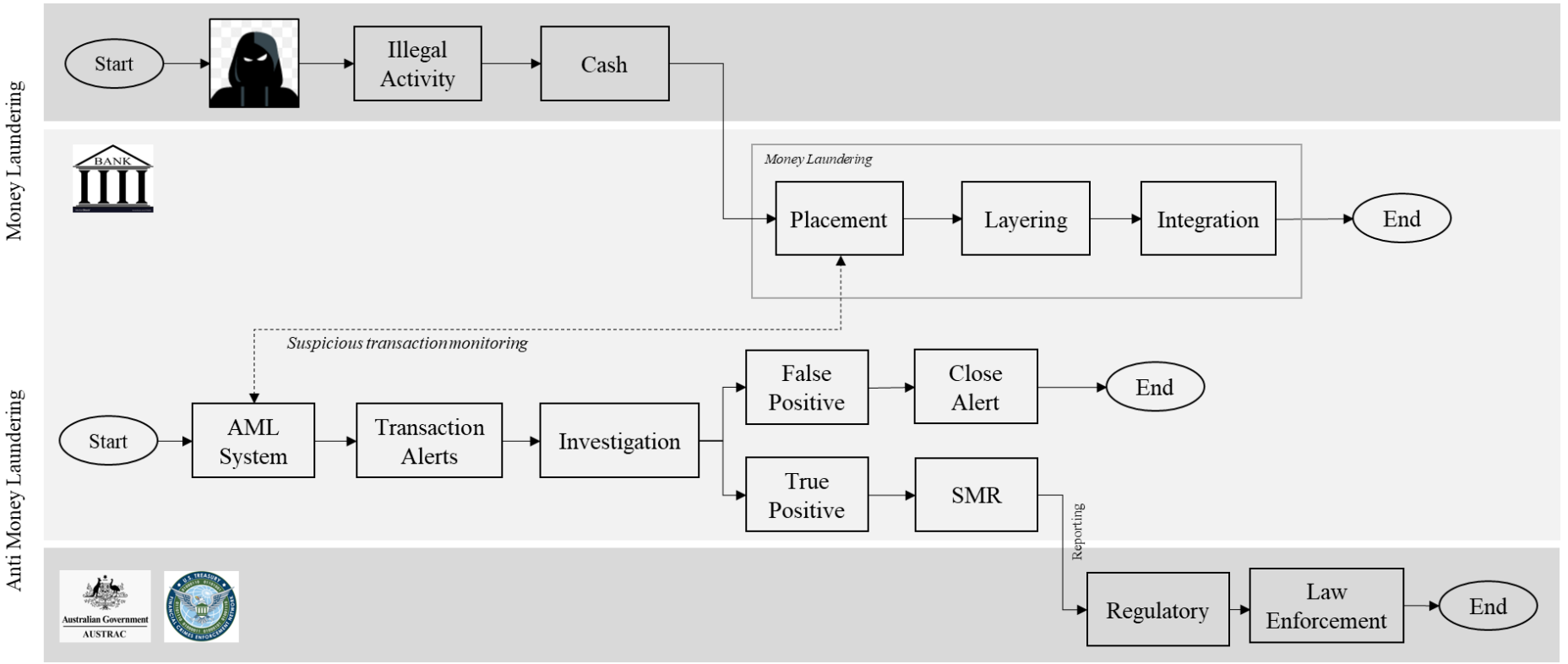


Figure 1.2 – Money laundering and AML process

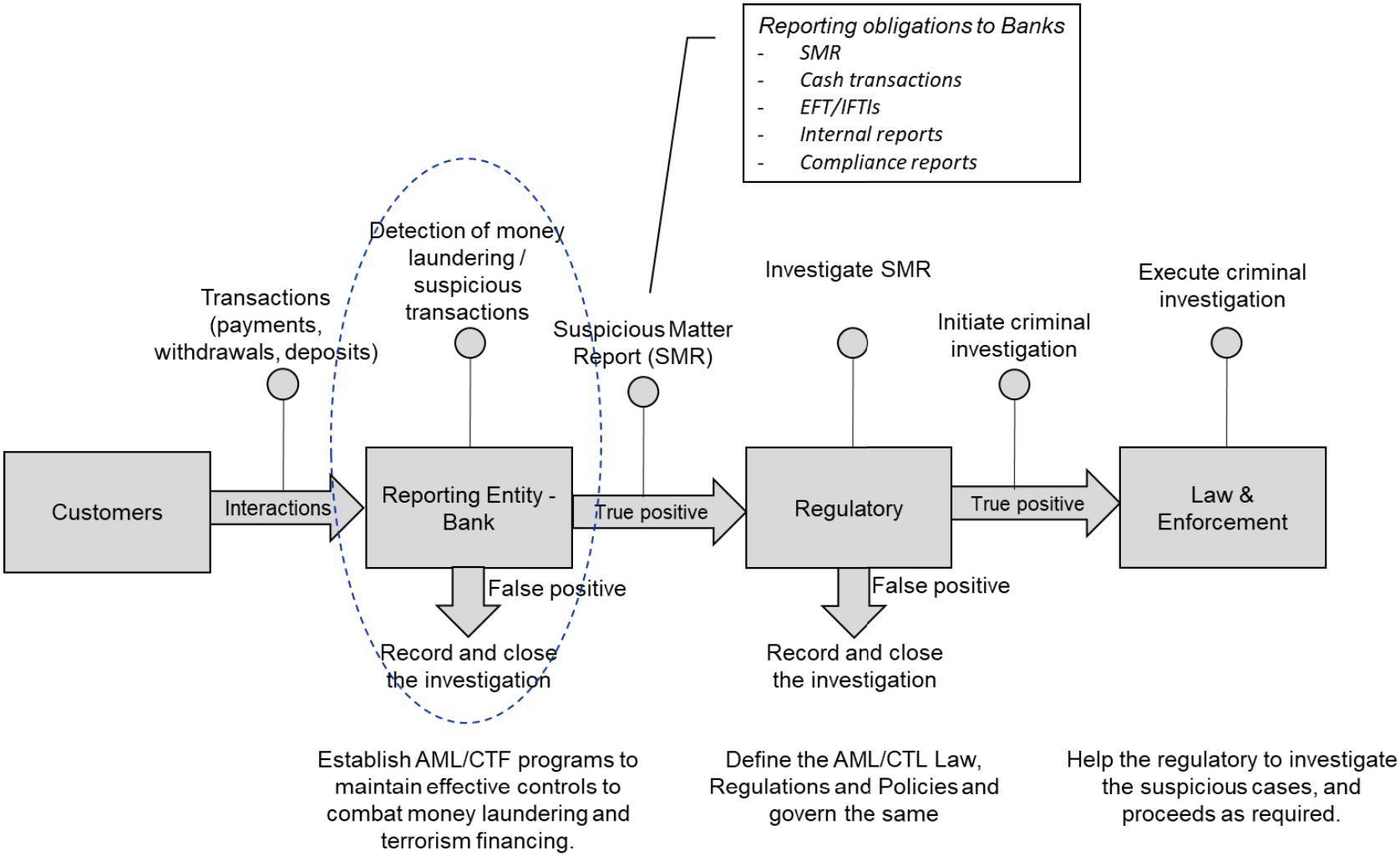


Figure 1.3 – AML ecosystem

1.3 Research Gaps

Due to recent advancement in communication and technology, many banking services are made available to customers through different online channels. This has caused an explosion in a number of daily transactions. Due to increase in transactions and continuously changing regulatory landscape, staying ahead of finance crime and compliance risk has become more complex and expensive than ever before for financial institutions.

By realising the complexity of money laundering domain, volume of the transactions and the success of AI in other fields to improve operational efficiency and prediction accuracy, few banks have started putting in the efforts to automate the data and time-intensive tasks using Artificial Intelligence / Machine Learning (AI/ML) technologies (Finance 2018) for detecting suspicious transactions. However, the adoption of machine learning is facing several challenges such as high false negative rate by shallow learning methods, lack of transparency of the models and explainability of the decisions made by opaque models. There is always a trade-off between model accuracy and interpretability while choosing a machine learning model. The black-box models such as neural networks, gradient boosting models (Farrugia et al. 2020) can predict highly accurate results, however, often lags on interpretability, hence it is difficult to explain the rationale behind the decisions made. On the other hand, the white-box models such as decision tree (Liu et al. 2011), linear regression are highly interpretable and easier to explain the decisions made by models, but provide less accurate results (Chen et al. 2018b). Usually black-box models are chosen by considering the high accuracy expectation in suspicious transaction detection, however, the lack of transparency on how decisions are made by these models to conclude a transaction is suspicious; and a mandatory ask by regulators to provide rationales behind suspiciousness of transaction, is posing an impediment for adoption of AI/ML technology in finance crime units in financial institutes (Barredo Arrieta et al. 2020).

XAI is a class of system that provides visibility into how an AI system makes the decisions, predictions and executes its actions (Rai 2020). In most recent years, tools like Local Interpretable Model-Agnostic Explanations (LIME) and SHAP are developed to

provide explanations as a Post-Hoc Analysis for complex AI/ML models, but still remains a long way until a general framework for explainability is established. This absence of explainable AI/ML systems in the AML domain, therefore, translates to an urgent demand for research into explainable systems for delivering high prediction accuracy without sacrificing explainability.

1.4 Problem Statement

Banks are using rule-based systems (primarily driven by threshold values) to identify suspicious transactions to report it to regulatory as a compliance obligation. It suffers from the following problems.

- Difficulties in keeping the rule-based AML systems up to date together with changing fraud patterns and regulatory landscape (Jullum et al. 2020).
- High false positive alerts raised for suspicious transactions by existing AML solutions (Davenport 2020), causes an increase in operational cost and lowers the operational efficiencies.
- Lack of Explainability by highly accurate machine learning techniques, causes a barriers to adoption of the technology (PricewaterhouseCoopers 2017).

1.5 Research Questions

Deep learning has shown good results in other domains such as computer vision (Sabokrou et al. 2021), medical image analysis (Palraj and Kalaivani 2021), speech recognition (Li et al. 2021) and natural language processing (Roshanzamir et al. 2021), however, the usage in AML domain is very limited. The research proposes to address the following questions:

- What deep learning methods are proposed to identify suspicious money laundering transactions?
- What type of data (synthetic or real data) is used to train, test and validate the proposed money laundering detection methods?

- What methods are interpretable among the ones that are proposed for suspicious transaction detection?
- Which deep learning classifiers can help to improve the accuracy in detecting suspicious money laundering transactions in financial institutions?
- What XAI techniques are proposed to explain the decisions made by deep learning methods? Is XAI used in money laundering domain?
- Which XAI technique is more effective to bring out the human interpretable explanation from the chosen deep learning classifiers in the above question?

1.6 Research Aim and Objectives

The primary objective of the research is to design a method that can detect the suspicious money laundering transactions most accurately along with the explanations. The sub-objectives of the research are as follows:

- **Objective 1** - Design and construct an effective deep learning technique to reduce false negative rate in detecting the suspicious money laundering transactions.
- **Objective 2** – Formulate an effective XAI technique to produce human interpretable explanation for the AML decisions made by the deep learning method chosen in objective 1.

1.7 Scope of Thesis

This research aims to develop a method to detect suspicious transactions using deep learning classifier, and a method to explain the decisions made by deep learning classifiers using state-of-the-art XAI techniques.

Following is a high level scope of the thesis:

- Design, develop a utility to create synthetic finance data as below:
    - Retail banking customers with savings account and transactions.
    - Suspicious money laundering transactions as a labelled data.

- Design a CNN model to detect suspicious transactions.
- Design a method to apply SHAP XAI technique on CNN.
- Design the following most frequently used machine learning models to detect suspicious transactions and compare the CNN results.
    - RF
    - SVM
    - XGBoost
- Model training and fine tunning
    - Data pre-processing, adding noise
    - Train CNN model and fine tune the hyper parameters for optimal results.
    - Train RF, SVM and XGBoost model and compare CNN results.
    - Apply SHAP on CNN to interpret the decisions made by CNN.
- Discuss the findings, limitations and recommendations for future research.

1.8 Motivation behind Research

Money laundering is a criminal activity used to disguise the illegally obtained money and make them appear legitimate in the system. The financial systems are being exploited over 5 decades for money laundering. While financial institutions have established a separate operation unit called as Finance Crime and Fraud, whose core responsibility is to manage the risk for Bank by ensuring compliance with AML/CTF law, and prevent terrorism financing and money laundering. In spite of spending huge money on employing an army of investigation officers and niche software products, financial institutions are still heavily penalised for non-compliance with AML/CTF law. This shows a clear solution gap and an opportunity to solve the same.

On the other hand, as technology is advancing, the Banks are digitising their operations to improve the customer experience by enabling several digital channels to interact with Banks easily. This has made banking and transactions possible at a few clicks away, which leads to an enormous increase in number of transactions and frauds as well. Day by day, the transactions are growing at an exponential speed, new fraud patterns are being innovated by fraudsters, and the regulatory landscape is changing continuously which is making it difficult to keep the rule based AML systems up to date. The AML officers are

struggling for a time to investigate the huge number of alerts generated by AML systems. This drives a need to put the AI in place to solve this problem. While it is understood that AI cannot replace the AML officers, but definitely can analyse the data to identify suspicious transaction activities along with supporting evidences that makes the task of AML officer easy to make the decisions.

Given this context, author firmly believes that suspicious transaction detection is an apt use case where machine learning, deep learning and XAI techniques should be used to improve the quality of decision support system for AML officers to increase the operational efficiency and manage the compliance risk effectively; and in turn effectively contribute in reducing the finance crime and associated criminal activities to drive peace and harmony in the society.

1.9 Novelty and Main Contribution

Theoretical significance

In banks, the finance data (including transactions, customers, and accounts) is one of the highly sensitive data asset and with several restrictions for consumption even within the bank. The finance data availability for research purpose is very scarce, and hence the research in AML domain is limited to only a few machine learning techniques and negligible research in deep learning space.

As a part of this thesis, the author has generated the synthetic data for retail banking customers which is close to real-time transactions and injected the synthetic money laundering transactions into the data. This dataset is used as a training and testing dataset for the deep learning classifier – CNN. To measure the performance of CNN, 3 machine learning classifiers (RF, XGBoost and SVM) are trained and tested on the same datasets. Further, the SHAP XAI method is applied on CNN model to derive the explanations of the predictions made by CNN model.

Application of deep learning method (CNN) to detect the suspicious money laundering transactions in banking domain and post-hoc explanation technique (using SHAP) to describe the predictions made by deep learning method is an addition to scientific knowledge base. This is a first time CNN together with SHAP used in the AML domain.

Practical significance

Financial institutions employ multiple solutions to combat money laundering such as following KYC, Customer Due Diligence (CDD), risk profiling, watchlist screening, sanctions, employee training programs, and transaction monitoring for suspicious activities. Considering the continuous increase in transactions, changing fraud patterns and regulatory landscape, a robust end-to-end framework utilising AI/ML would significantly help the financial institutions as below:

- It would improve the effectiveness of detecting suspicious money laundering transactions that leads to reduction of compliance risk and operational cost required for screening and investigation of the false positive alerts (Finance 2018), (Al-Shabandar et al. 2019).
- The explanations generated by post-hoc methods for machine learning decisions would help AML officers to view the reasoning behind the predictions and gather the relevant evidences to enhance the investigation, and report the suspicious transactions with sufficient evidences and high confidence.
- Explainable AML system would help to build the trust (PricewaterhouseCoopers 2017) and drive the adoption of AI/ML technologies in AML domain which can help to manage the compliance risk, keep the brand reputation intact and avoid hefty penalties for non-compliance.

1.10 Thesis Organisation

The thesis is structured in 5 key chapters.

Chapter 1 provides thesis introduction, background on money laundering domain, research context, research gaps, problem statement, research questions, research objectives, scope, research motivation and significance.

Chapter 2 provides the literature review of machine learning techniques used in money laundering detection. It includes review methodology, review questions, review outcomes including machine learning methods, deep learning methods, XAI methods; discussion

including limitations of existing research, current research challenges, barriers for adoption of machine learning technologies in the finance crime domain, future research directions and summary of the review.

Chapter 3 discusses the core research methodology. This includes a synthetic data creation approach covering data attributes, data creation rules, money laundering scenarios to train the models; design of CNN classifier to detect suspicious transactions, design of traditional machine learning methods including RF, SVM, XGB to compare the CNN results; design of SHAP XAI technique to explain the decisions made by CNN classifier; software tools used for modelling and synthetic data creation and finally presents the summary of the chapter.

Chapter 4 discusses the key findings and results produced by conducting several experiments to address the research objectives. It includes synthetic data generated, modelling results covering suspicious transaction predictions using CNN, RF, SVM and XGBoost; comparison of CNN results with RF, SVM and XGBoost using the performance metrics chosen for research; the outcome of applying SHAP to explain the CNN predictions; and finally a summary of the chapter.

Chapter 5 provides the conclusion that includes the key findings of the research to address objective 1 and objective 2, practical implications from bank perspective, research drawbacks and limitations, recommendations for future research direction.

2 LITERATURE REVIEW

This section provides the literature review to find the current state of research in application of machine learning, deep learning and XAI techniques in the AML domain, analyse the published methods and identify the research gaps with respect to current need.

2.1 Introduction

Banks are obliged to monitor the transactions to detect the suspicious activities, investigate it, and report the same to regulatory to ensure compliance with AML/CTF policies defined by government. Effective AML systems, controls, practices are critical to manage the compliance risk. **Figure 2.1** presents a conceptual process followed in the financial institutions for transaction monitoring, suspicious transaction detection and reporting the same to the regulatory.

The decision of transaction being suspicious goes through 5 steps:

[1] Alert generation system (mostly automation) – AML system monitors the transactions for any suspicious activities and the transactions that are beyond certain thresholds which are required to be reported to regulatory, and generate the alerts.

[2] Manual investigation at financial institution – AML officers investigate these alerts and identify if they are genuine alerts or false alerts. In case of a genuine alert, a case is filed, else the alerts are ignored.

[3] Case investigation and SMR preparation – the case is further investigated by AML domain experts to gather adequate evidences of suspiciousness. If the evidences are not strong enough, then a case is closed, else a SMR report is prepared and sent to regulatory (or FIU).

[4] Investigation at FIUs – the SMRs are investigated by FIU staff in a larger context and investigate if that needs criminal investigation and if required, the same is handed over to Law & Enforcement departments else the SMR is closed.

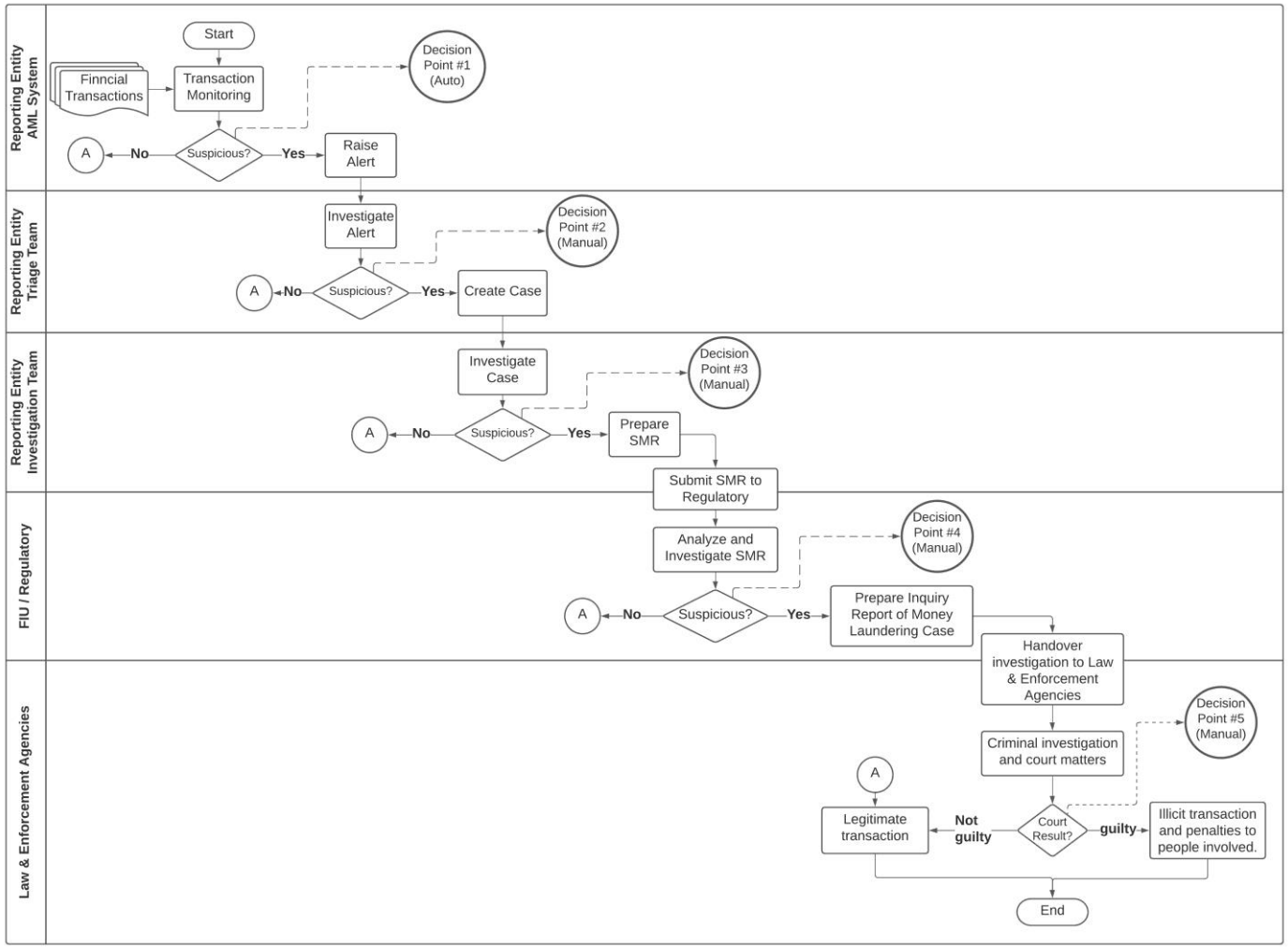


Figure 2.1 – AML Process

[5] The proceeds at court – criminal investigation report and court matter decides if it is a genuine case of money laundering involving criminal activities or not and accordingly the case gets closed.

It is evident that the suspicious transaction detection is a multi-step classification problem, because the confidence level of transaction being suspicious is built upon the investigation performed at each step from Alert to SMR and beyond the purview of financial institution. Considering the efforts involved by various parties (Bank, FIU and Law & Enforcement agencies), it is important to get the detection of suspicious transaction correct to reduce the efforts.

2.2 Literature Review Methodology

The study followed a standard 3 step review process described by Kitchenham (Kitchenham and Charters 2007) – Plan, Conduct and Report. **Figure 2.2** presents the methodology in detail.

In plan stage, inputs and triggers are gathered from a study of current challenges in finance crime domain, where it is found that detection of money laundering is a major issue in Banks. That is evident from the fact that several financial institutions are penalised for non-compliance with AML and terrorism financing policies set by government (Centre 2020),(Centre 2018). Based on the literature review, the aim and questions to be addressed were defined.

In conduct stage, we decided to find the existing work done in the space of application of machine learning, deep learning and XAI technique in AML domain. For this a literature search is done using the combinations of the following keywords - "*money laundering*", "*suspicious transaction*", "*anti-money laundering*", "*deep learning*", "*machine learning*", "*data mining*", "*statist**", "*explainable AI*" and "*XAI*". The papers were segregated based on the relevance to main topic and most relevant papers considered for review. The review focused on identifying the application of deep learning techniques for money laundering detection, reviewing the methods including machine learning to know if they are interpretable or not, identify if any XAI techniques are applied where non-interpretable methods are used and present the type of data used during model training and evaluations.

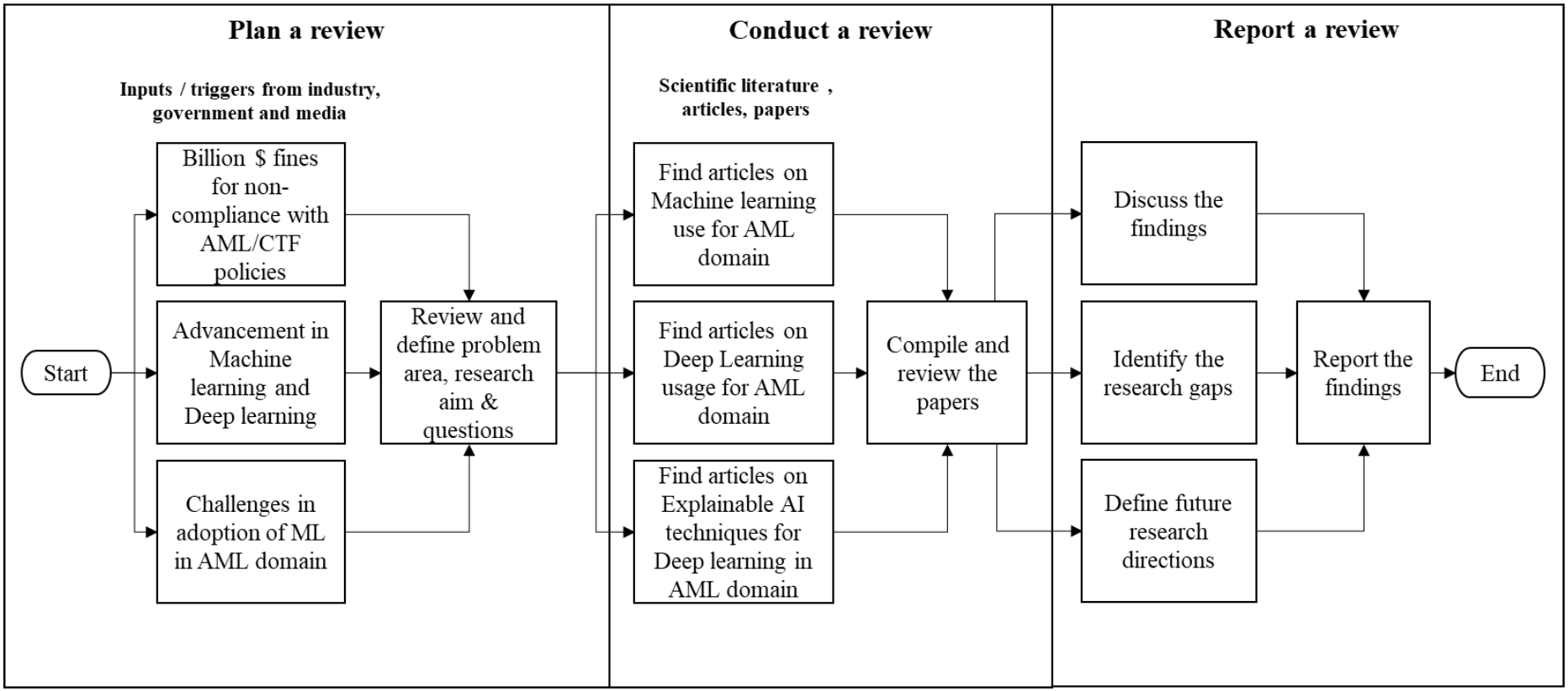


Figure 2.2 – Literature review methodology

Study also brought in the viewpoint of key review papers on “Machine Learning techniques on AML” and “XAI techniques” subject from literature.

In the report stage, the study focused on analysis of the findings, performed assessment and reported the key challenges and constraints for researchers, barriers for AI/ML adoption in AML domain, future research directions.

2.2.1 Aim and Research Questions

The key motivation for this review is to identify the existing literature on money laundering detection that shows the usage of statistical methods, data mining and machine learning techniques and deep learning methods to assess the current state of research. The interpretability of the machine learning classifiers is important from AML domain perspective but there isn't any comprehensive study that provides the details on current state of interpretability of the research published so far and the application of state-of-the-art XAI techniques in AML domain.

Hence, the objectives of this review are - to identify the deep learning techniques applied in money laundering domain, determine the interpretability of the research published so far, identify the XAI techniques applied in AML domain to bring out the explainability; and answer the following questions:

- Is there an adequate research in the space of money laundering detection using machine learning, data mining? What methods are proposed?
- What deep learning methods are proposed to identify suspicious money laundering transactions?
- What XAI techniques are proposed to explain the decisions made by deep learning methods? Is XAI used in money laundering domain?
- What type of data (synthetic or real data) is used to train, test and validate the proposed money laundering detection methods?
- What methods are interpretable among the ones that are proposed for suspicious transaction detection?

2.2.2 Search and Selection Process

Table 2.1 describes the combinations of keywords (money laundering, anti-money laundering, suspicious transaction, explainable AI, XAI, machine learning, deep learning, statist*) used for searching papers from Scopus library.

Following search criteria used while searching the articles on the Scopus library.

- The keyword search criteria - “Search within: Article title, Abstract, Keywords”.
- Document type filter – Included articles from the following category – Article, Conference Paper, Review, Book Chapter, Conference Review, Book, Data Paper, and Short Survey. Articles from the Retracted category are excluded.
- Publication stage filter - articles with publication stage as Final are considered, whereas articles with Article in the press stage are excluded.

Table 2.1 shows the search query and raw result of the queries on the Scopus library. These queries are designed to search the research articles in the money laundering detection domain that applies the following technical domain – XAI, machine learning, deep learning, data mining, and statistical methods.

Table 2.1 – Literature search queries and results

No.	Search Query	Papers
1	("money laundering" OR "suspicious transaction" OR "anti-money laundering") AND ("explainable AI" OR "XAI")	4
2	("money laundering" OR "suspicious transaction" OR "anti-money laundering") AND ("machine learning")	114
3	("money laundering" OR "suspicious transaction" OR "anti-money laundering") AND ("deep learning")	23
4	("money laundering" OR "suspicious transaction" OR "anti-money laundering") AND ("data mining")	112
5	("money laundering" OR "suspicious transaction" OR "anti-money laundering") AND (statist*)	148

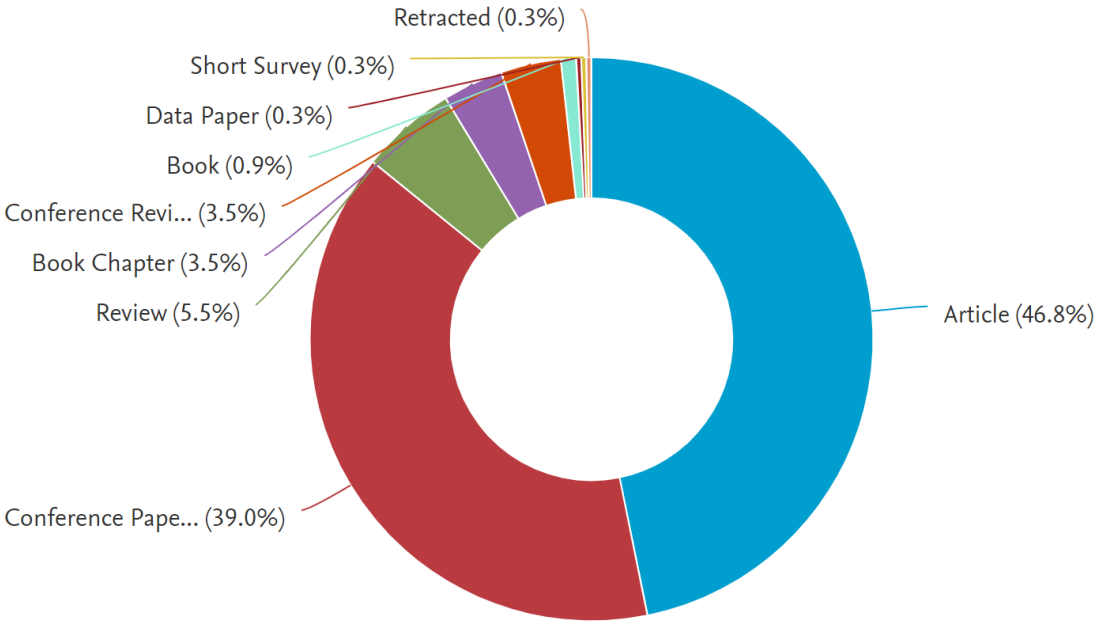


Figure 2.3 – Literature analysis – by article type

Figure 2.3 shows the combined list of articles by type, found in literature as per the search queries listed in **Table 2.1**.

Figure 2.4 shows the list of articles published year wise on the money laundering detection domain and highest number of articles are seen published in year 2021. It is evident that the research is getting more attention year by year.

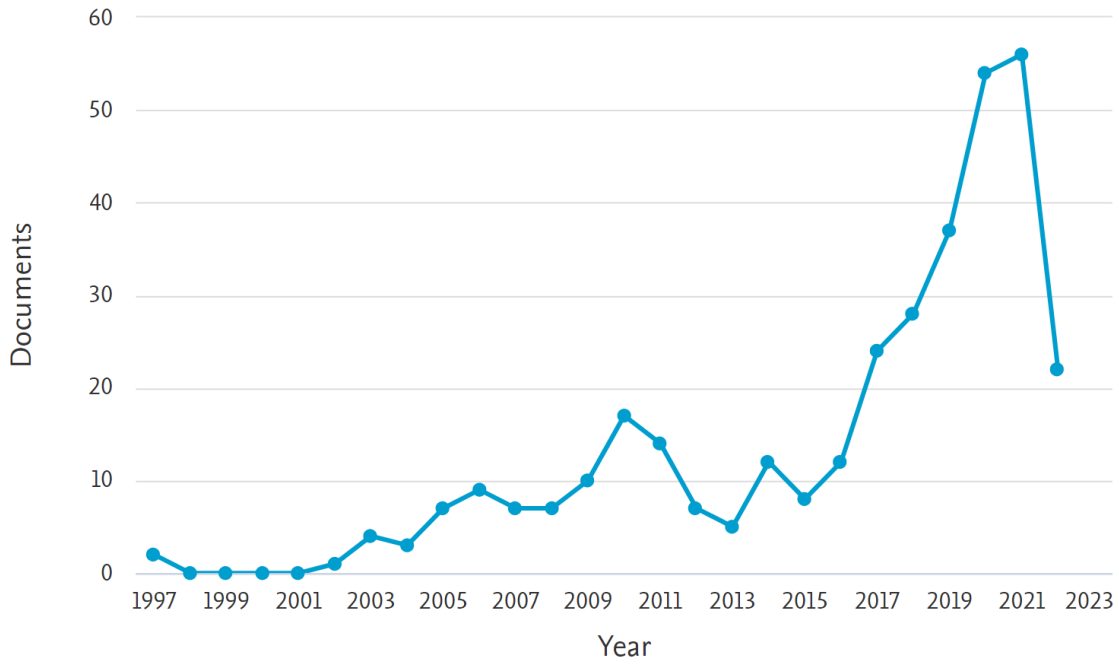


Figure 2.4 – Literature analysis – year wise articles

Figure 2.5 shows the analysis of the journals where the money laundering research articles are published most and also show the year wise numbers for the journals.

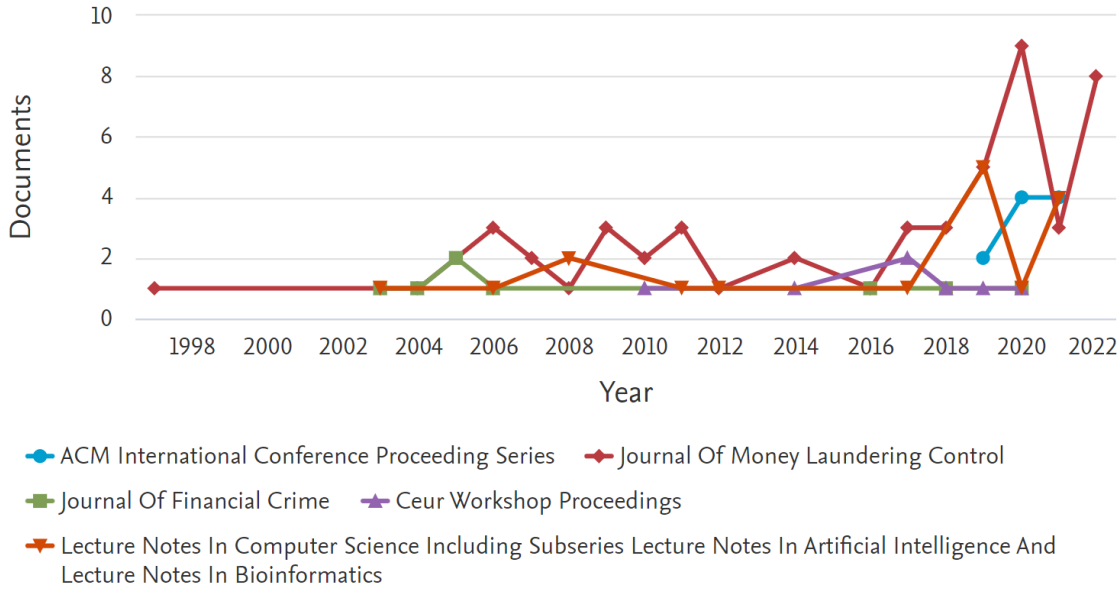


Figure 2.5 – Literature analysis – articles by journals

After skimming through 345 articles found in search results, the duplicate and non-relevant papers are excluded from the analysis. The refined list of papers are classified in 4 categories as shown in **Table 2.2**.

Table 2.2 – Literature classification

Literature classification	Statistical methods & Data mining	Machine learning	Deep learning	XAI methods
Methods in AML	37	37	5	-
Review papers	8	8	-	3

The papers included for further analysis are:

- Around 80 papers found which are directly related to money laundering. It includes detection of suspicious money laundering transactions, patterns, and groups. Hence, they are considered for further review.
- Around 19 review papers found from machine learning, data mining, XAI hence they are used for further analysis.
- Papers from the finance crime domain in general are used to build the subject context.

The articles excluded for further analysis are:

- Papers found from fraud domain, but not related to the money laundering domain are excluded.
- Non-relevant papers to the money laundering topic are excluded.
- Duplicate papers in the search results are excluded

In addition to the paper found from a systematic search, the information from the following category of literature was also used to establish the context. These papers are searched separately through independent sources including UTS library, internet and websites specific to organizations.

- Regulatory and industry – policy documents describing the compliance requirement, typologies and case study reports describing money laundering cases.

- Penalties for non-compliance – media newspaper, respective financial institute's websites, regulatory websites. This source of information indicates the effectiveness of the current solutions being used to fight money laundering without getting much into details explicitly.

2.3 Various Techniques in Anti-Money Laundering

In financial institutions, all transactions data coming from several internal banking systems such as retail banking, consumer banking, wealth management, institutional banking, etc.; flows through the rule-based AML system, where each transaction is assessed against the pre-defined rules to identify (a) the transactions of $10,000 or more, (b) money transferred to and from overseas, and (c) suspicious activities involving money laundering or financial frauds. While rule-based systems are essential in banks and work well for identifying transactions as per pre-defined rules (category 'a' and 'b'), but it struggles to detect the emerging fraud patterns (category 'c'). If the transaction matches with the rule, a red flag or alert is raised, which is processed and reported to the regulatory. The basis for rule definition is threshold amounts for different type of transactions based on the regulatory requirements (Network 2021; Office of Parliamentary Counsel 2019) and the recommendations given by the FATF (Force 2020).

The key issues with rule-based systems from 'suspicious transaction detection perspective' are, keeping the rules up-to-date and relevant at all the times, and weighing different rules is almost impossible (Jullum et al. 2020), generation of a large number of false positive alerts – up to 98% (Davenport 2020) and inability to handle a high volume of different type of structured data, semi-structured data or unstructured data (Chen et al. 2018c). More alerts means more efforts by the compliance officer to scan through it. If the officer confirms that the transaction can be justified enough to be true suspicious, then such transactions are required to be reported to regulatory and it is an obligation for financial institutions to do so (Network 2021; Office of Parliamentary Counsel 2019). The decision of concluding a transaction as suspicious is very important decision, if it goes incorrect the bank may end up reporting a good customer to regulatory for further investigation which may result in a criminal investigation by authorities. On the other

hand, not identifying bad customers may result in continuing the money laundering activities by leveraging the banking systems and more finance crime.

Considering the limitations of rule-based systems, researchers have started leveraging the machine learning technologies to automatically identify the suspicious transaction patterns using a variety of solution categories such as AML typologies, link analysis, behavioural modelling, risk scoring, anomaly detection, and geographic capability (Chen et al. 2018c).

2.3.1 Machine Learning Techniques

There has been a significant research in the space of applying machine learning methods to detect money laundering such as identification of suspicious transactions, money laundering gangs and laundering patterns. **Table 2.3** presents the machine learning methods used for the detection of money laundering.

Table 2.3 – Machine learning methods used in AML domain

No	Method	Year	Reference
1	Decision tree	2021	(Baesens et al. 2021b)
		2021	(Tundis et al. 2021)
		2021	(Alkhalili et al. 2021)
		2021	(Rocha-Salazar et al. 2021)
		2020	(Garcia-Bedoya et al. 2020)
		2019	(Zhang and Trubey 2019)
		2018	(Badal-Valero et al. 2018a)
		2016	(Alexandre and Balsa 2016)
		2011	(Liu et al. 2011)
		2010	(George and Kavakli 2010)
		2008	(Zhang and Zhao 2008)
2	RF	2022	(Labanca et al. 2022)
		2021	(Tundis et al. 2021)
		2020	("ICMLT 2020 - Proceedings of 2020 5th International Conference on Machine Learning Technologies" 2020)
		2020	(Garcia-Bedoya et al. 2020)
		2019	(Zhang and Trubey 2019)
		2019	(Baek et al. 2019)
		2018	(Magomedov et al. 2018)
		2018	(Badal-Valero et al. 2018a)
		2017	(Savage et al. 2017a)
3	SVM	2021	(Alkhalili et al. 2021)
		2021	(Tundis et al. 2021)
		2019	("16th Web Information Systems and Applications Conference, WISA 2019" 2019)
		2019	(Zhang and Trubey 2019)
		2017	(Savage et al. 2017a)
4	XGBoost	2021	(Baesens et al. 2021b)
		2020	(Tertychnyi et al. 2020)

No	Method	Year	Reference
		2020	(Farrugia et al. 2020)
		2020	(Garcia-Bedoya et al. 2020)
		2020	(Jullum et al. 2020)
5	Clustering	2021	(Rocha-Salazar et al. 2021)
		2019	(Chao et al. 2019)
		2014	(Yang et al. 2014)
		2012	(Cao and Do 2012)
		2011	(Raza and Haider 2011)
		2010	(Le Khac and Kechadi 2010)
6	Artificial neural network	2019	(Zhang and Trubey 2019)
		2018	(Badal-Valero et al. 2018a)
		2010	("Digital Forensics and Cyber Crime - First International ICST Conference, ICDF2C 2009, Revised Selected Papers" 2010)
		2010	(Le Khac and Kechadi 2010)
		2007	(Yang et al. 2007)
7	Logistic regression	2021	(Baesens et al. 2021b)
		2020	(Tertychnyi et al. 2020)
		2020	(Garcia-Bedoya et al. 2020)
		2018	(Badal-Valero et al. 2018a)
8	Graph learning	2019	(Kanezashi et al. 2019)
		2018	("Proceedings - 2018 IEEE 6th International Conference on Future Internet of Things and Cloud Workshops, W-FiCloud 2018" 2018)
		2011	(Michalak and Korczak 2011)
9	Isolation forest one class SVM	2020	(Guevara et al. 2020a)
		2020	("Proceedings of the 13th IADIS International Conference ICT, Society and Human Beings 2020, ICT 2020 and Proceedings of the 6th IADIS International Conference Connected Smart Cities 2020, CSC 2020 and Proceedings of the 17th IADIS

No	Method	Year	Reference
			International Conference Web Based Communities and Social Media 2020, WBC 2020 - Part of the 14th Multi Conference on Computer Science and Information Systems, MCCSIS 2020" 2020)
		2017	(Camino et al. 2017)
10	Naïve Bayes	2021	(Alkhalili et al. 2021)
		2021	(Tundis et al. 2021)
		2020	(Kumar et al. 2020)
11	Multi-layer perceptron	2020	("ICMLT 2020 - Proceedings of 2020 5th International Conference on Machine Learning Technologies" 2020)
		2019	("16th Web Information Systems and Applications Conference, WISA 2019" 2019)
12	AutoEncoder	2020	(Garcia-Bedoya et al. 2020)
		2019	(Paula et al. 2017b)
13	Linear regression	2021	(Tundis et al. 2021)
		2019	("16th Web Information Systems and Applications Conference, WISA 2019" 2019)
		2019	(Chao et al. 2019)
14	Link analysis	2014	(Yang et al. 2014)
		2003	(Zhang et al. 2003)
15	Dynamic Bayesian networks	2019	(Chao et al. 2019)
		2011	(Raza and Haider 2011)
16	Genetic algorithm	2010	(Le Khac and Kechadi 2010)
		2007	(Yang et al. 2007)
17	Expectancy maximization	2019	(Baek et al. 2019)
		2017	(Camino et al. 2017)
18	Extra trees	2020	("ICMLT 2020 - Proceedings of 2020 5th International Conference on Machine Learning Technologies" 2020)

No	Method	Year	Reference
19	Bagging classifier	2020	("ICMLT 2020 - Proceedings of 2020 5th International Conference on Machine Learning Technologies" 2020)
20	Graph convolutional network	2020	("ICMLT 2020 - Proceedings of 2020 5th International Conference on Machine Learning Technologies" 2020)
21	Bayes Logistic Regression	2019	(Zhang and Trubey 2019)
22	MinMaxScaler method	2018	(Magomedov et al. 2018)
23	Affiliation Mapping Calculation and Sequential Mining	2016	(Jayasree and Siva Balan 2016)
24	Probabilistic Relational Model using the Audit Sequential Pattern (PRM-ASP) Mining	2015	(Jayasree and Siva Balan 2015)
25	Predictive Security Analysis at Runtime (PSA@R)	2014	(Zhdanova et al. 2014)
26	Anomaly detection	2011	(Raza and Haider 2011)
27	K-means	2010	("Digital Forensics and Cyber Crime - First International ICST Conference, ICDF2C 2009, Revised Selected Papers" 2010)
28	Text mining	2007	(Yang et al. 2007)
29	Outlier detection	2020	(Torres and Ladeira 2020)
30	Gaussian Mixture Models	2017	(Camino et al. 2017)
31	Community detection using temporal-directed Louvain algorithm	2017	(Li et al. 2017)
32	Apriori, PrefixSpan, FP-growth and Eclat algorithms	2016	(Drezewski et al. 2016)
33	Scan statistics	2010	(Liu and Zhang 2010)
34	Semantic Core Tree	2006	(Chen et al. 2006)
35	Multi-channel Convolutional Network based on NLP	2015	(Han et al. 2015a)

While searching for literature on the application of machine learning techniques in the money laundering domain, we found a few comprehensive and systematic review papers published over the decade that describe data mining and machine learning methods applied in AML domain. This section will not describe those methods again here, instead list the most recent review papers in **Table 2.4** and briefly summarise the finding:

Table 2.4 – Review papers – machine learning in AML domain

No.	Literature reviews of AML	References
1	Review of machine learning methods	(Milind Tiwari 2020), (Chen et al. 2018c), (Labib et al. 2020), (Alsuwailem and Saudagar 2020)
2	Review of data mining methods	(Ngai et al. 2011)
3	Application of statistical methods	(Sudjianto et al. 2010)
4	Application of technological solutions	(Leite et al. 2019)

- **Review of machine learning methods used for money laundering detection**

A literature review of money laundering and its related area was conducted by (Milind Tiwari 2020), with the aim of identifying the gaps and directing attention towards addressing them. The key findings are categorized into six groups as – (i) AML framework and effectiveness, (ii) impact of money laundering on economy, (iii) money laundering ecosystem and motivation, (iv) magnitude of money laundered, (v) avenues for money laundering, and (vi) detection methods of money laundering. As per the findings by (Milind Tiwari 2020), most studies of detecting money laundering have focused on transactions in banks, real-estate and trade based companies; however, the literature on detection of shell companies used for money laundering is meagre. Shell companies established in the UK alone, were found to be linked with laundering £80bn between 2010 and 2014. The paper (Milind Tiwari 2020) has highlighted this gap in the literature and proposed a need for developing strategies to identify the shell companies involved in illicit activities.

A comprehensive survey conducted by (Chen et al. 2018c) presents machine learning algorithms and methods applied to detect suspicious transactions. The results are organized in 6 solutions categories – AML typologies, anomaly detection, behaviours modelling, link analysis, geographic capability and risk scoring. Various methods are analysed and compared. Authors have identified the key capabilities of the AML system as – efficient data preparation, data transformation and data analytics techniques. The key findings from the review is – inadequate focus on data quality assurance in the published research. Future research direction is to employ reinforcement learning to train the model considering the agility of financial operations.

A review conducted by (Labib et al. 2020) has brought out the machine learning algorithms used for identifying the money laundering patterns, detect unusual behaviour, identify money laundering groups, and detecting money laundering groups. The identified machine learning algorithms/techniques are – Rule Based Methods (Senator et al. 1995),(Wang and Yang 2007), Decision Trees (Zhang and Zhou 2004), Artificial Neural Networks (ANN), SVM, RF, Outlier Detection Methods, Social Network Analysis (SNA), Naive Bayes, K-Nearest Neighbour (KNN), deep learning, graph mining, K-Means Clustering and One Class SVM. These identified methods have been thoroughly studied and compared to present how they behave for large and highly imbalanced datasets. The key findings of these papers are - each method has some strength and weakness, hence, the method needs to be leveraged as per the circumstances, though supervised and semi-supervised machine learning showed decent results and considered as out-of-the box solutions however they cannot find new machine learning patterns and have a high rate of false positives, therefore author feels unsupervised machine learning is the way forward to detect new money laundering patterns.

A literature review conducted by (Alsuwailem and Saudagar 2020) focuses on the papers published between 2015 to 2020 to understand the state-of-the-art in AML systems, presents the results using the following categories – supervised learning, unsupervised learning, data sources, evaluation methods, implementation tools, sampling techniques and regions of study. The key findings by (Alsuwailem and Saudagar 2020) are – Decision Tree, Radom Forest and SVM are most frequently used algorithms in AML system from supervised category, neural networks is mostly used in unsupervised

category; Accuracy, Area Under the Curve (AUC) and precision are used for model evaluation; most of the data used for research was customer and transaction data from the banks. However, there are many other methods for laundering the money such as restaurants, hotels and law offices, which are not researched enough.

- **A review of data mining techniques in financial fraud detection**

A comprehensive literature review was conducted by (Ngai et al. 2011) on application and classification of data mining techniques for financial fraud detection by critically reviewing 49 journal articles published between 1997 and 2008. The financial fraud is classified into four categories as – bank fraud, insurance fraud, securities, and commodities fraud, and other relevant financial frauds. The data mining techniques are classified into six categories – (i) classification, (ii) regression, (iii) prediction, (iv) outlier detection, (v) visualization, and (vi) clustering. The key findings are – data mining techniques are largely applied in detection of insurance frauds, followed by corporate and credit card fraud, and identified a need of applying these techniques from money laundering, mortgage fraud and securities fraud. Only one technique 'network analysis' is found for detecting of money laundering. Otherwise, largely Logistic Models, Neural Network, Bayesian Neural Network, and Decision Trees are found to be used for detection of overall frauds.

- **A survey of statistical methods for detection of money laundering**

The survey conducted by (Sudjianto et al. 2010) provides a list of statistical methods, data mining and machine learning techniques used for detection of money laundering and retail banking frauds. The authors identified the key challenges as – data volume and complexity, class imbalance, frequent evolving patterns, class overlap and, class mislabelling. The reviewed methods are classified in two categories as – supervised learning containing profiling, classification (SVM, BBN, HMM, neural network, classification Tree and classification rules) and link analysis; and unsupervised learning containing clustering, low-dimension representation and scoring, and anomaly detection. The key finding is, some methods are used effectively; however, it is also emphasized to

put more focus on preventing the fraud, example – employee training and awareness, defining policies and procedures.

- **A review of technological solutions for AML**

A systematic literature review conducted by (Leite et al. 2019), on application of technological solutions to combat money laundering by reviewing 71 papers published between 1997 and 2019. The results are presented using the following categories – domains of application approaches and classification of support systems. Application domains are defined as - suspicious transaction detection, pattern detection/groups/money laundering anomalies, risk assessment/analysis, security, control, structuring and/or governance applications and visual analysis/applications of visual techniques. Support mechanism categories are defined as – systems / software/ tools/ programming languages, hardware, patterns/ theories/ frameworks, algorithms/ mathematical application (data mining and machine learning). The key findings by (Leite et al. 2019) are, among the application domains, detection of suspicious transaction attracted more attention of researchers. From support systems adoption point of view, data mining techniques were used seldom for money laundering detection. A key gap identified from the papers was a need for data analysis description and evidence to support the benefits presented.

Apart from the reviews listed in the above 4 categories, (Irofti et al. 2021) presents the review of anomaly detection methods for fraud detection using graph structures, (Semenov et al. 2017) presents the analysis of anomaly detection, machine learning and neural networks, and (Shaju and Valliammal 2019) presents the measures for fraud detection.

It is evident from the literature described above that statistical methods and several machine learning techniques have been used for detection of money laundering. However, elaboration of deep learning technique usage is very limited. The following section describes the deep learning techniques used for money laundering as per the literature.

2.3.2 Deep Learning Techniques

Deep learning techniques have been used in various fields for predictions and found to be giving highly accurate results (Jaiswal and Valstar 2016; Liang et al. 2016; Szegedy et al. 2015). In finance fraud domain adversarial AutoEncoder together with machine learning classifiers is used by (Silva et al. 2021) for classifying the transactions as regular, local and global anomalies in mobile money service product. There were some challenges with deep learning such as a need for heavy computing, and more data. However, in today's world these challenges are no longer a big issue. This section describes the existing literature focused on the usage of deep learning methods for detecting money laundering.

- **AutoEncoder, Variational AutoEncoder and Generative Adversarial Network**

Supervised and unsupervised machine learning has their own merits and drawbacks. For example, supervised machine learning model capabilities are limited by the labelled data for training and generally not so good at identifying new patterns. Whereas unsupervised machine learning models are good at identifying anomalies and detecting new patterns. Weber et al. (Chen et al. 2021) has proposed a combination of unsupervised and deep learning classifiers to detect money laundering. Three deep learning classifiers - AutoEncoder (AE), Variational AutoEncoder (VAE), and Generative Adversarial Network (GAN) are used to design a method. Experiments conducted by the method demonstrates substantial reduction of false positives and improvement in accuracy.

- **Scalable graph CNN**

A scalable graph CNN is used for forensic analysis for financial data which is massive, dense and dynamic in nature (Mark Weber 2018). The outcome of the analysis in visual form is presented as one of the effective decision support for the AML analysts who are involved in a review of large amount of alerts raised by rule-based AML systems. The method was developed and evaluated using the synthetic graph (1M nodes and 9 M edges) generated using the AMLSim data simulator tool. The vertex represents the account with attributes such as account number, account type, owner name and account creation date. The edge represents the transaction with attributes such as transaction ID, amount, timestamp. Using escalated alerts and SMRs as a label data, a semi-supervised learning

model predicts the suspiciousness of a given node and potential bad actors in the transaction network based on the direct or indirect connections with the node. Author ran the experiment on a workstation with two deep graph model. First using the graph model GCN developed by (Kipf and Welling 2017), which took 611 seconds to converge in 32 epochs. Second experiment using the improved graph model FastGCN (Chen et al. 2018a), which took 386 seconds to converge on the same number of epochs. FastGCN is twice as fast as GCN. Considering the huge amount of financial data that goes as millions of transactions per second, the training speed is important to timely identify the suspicious activities. The key findings of this study (Mark Weber 2018) are - graph deep learning has also been emerging as an important tool for AML, due to nature of the domain where relationships can be easily established using account data as node and transaction data as edges, and the model is able to handle the large volume of data; however, author recommends further improvements to reduce the training time.

- **Multi-channel CNN**

A novel distributed and scalable framework is developed using deep learning driven natural language processing (NLP) technology to augment AML monitoring and investigation (Han et al. 2015b). The proposed framework performs different level of sentiment analysis, entity identification, relationship extraction, and link analysis on different data sources such as news, tweets, social media, etc. Each NLP module uses a specific data source to perform the analysis and bring the recommendations. The proposed framework is based on micro-service oriented distributed architecture and Advanced Message Queuing Protocol (AMQP) based integration platform, database as Cassandra, Neo4j and MySQL along with twitter and news engine. Two types of data are considered – financial data that includes KYC, Customer, Account, Transactions; and open data that includes financial news articles, financial reports, fraud open databases and social media contents. The system starts with identifying the suspicious transactions, then the key attributes are extracted from the alerted transactions such as name, location, account details which are used by next NLP modules. Finally, the suspicious transaction confidence score (in a range of 0.00 to 1.00) is calculated and presented by using the evidences produced by the following modules – TM, name screening, fraud knowledge base, sentiment analysis trend and entity disambiguation module. The sentiment analysis

is used to identify the polarity of the sentiment – either positive or negative. Author used multi-channel CNN based sentiment classifier to process financial news and CNN based sentiment classifier for social media data. Relation extraction is used to predict the relation between pairs of entities in the sentence. Document level SA classifier, trained on auto labelled 12,467 financial news articles, achieved 76.96 % accuracy. Author used a pipeline based method for relationship extraction tasks. Name Entity Recognition (NER) is performed using a combination of two methods – Stanford NER recognizer and neural NRE based on LSTM-CRF framework. The NER model was evaluated on SemEval 1010 task 8 data and reported 80.62% micro-f1 measure. For handling multi-instance problem author has developed an RNN with word-level and sentence-level attention for relation prediction. The RNN model reported 88% accuracy using P@100 measure post evaluation on New York Times dataset. Overall experiment is performed on synthetic and real world dataset. The results are verified by AML experts who mentioned that approximately 30% cost can be saved over the existing manual approach of AML investigation. To give a perspective of the manual efforts, authors have given one example of bank, where 10,000 analysts globally investigated over 400,000 alerts of suspicious transactions per week. As a manual investigating process, the analysts are required to investigate several sources of data that include news search, name screening, querying fraud databases, crime records, fraud offenses or any existing suspicious activities. Author has identified the improvement areas that can be considered for future research as – consider domain specific data and fine-tune the parameters, scale and deploy system on cloud for processing large data in real time, tailor the solution and evaluate it in other domains.

- **AutoEncoder**

The unsupervised deep learning based technique was proposed by (Paula et al. 2017a) to classify Brazilian exporters to find out the possibility of committing the frauds in exports. The model uses the AutoEncoder classifier to detect anomalies considering the regular transaction patterns in the data. Brazilian exports are reaching to ~200 countries with the help of 50,000 legal entities involved in shipping the goods. Author uses the database of foreign trade of the secretariat of federal revenue of Brazil for applying the deep learning model and finding out the export organizations whose explanatory variables of their

export operations show signs of divergence compared to regular pattern. The paper has followed the CRISP-DM methodology for implementation, considering the flexibility of the model. Two methods are evaluated – AutoEncoder and PCA using 819,990 records and, H2O software and H2O R package. From data preparation perspective, 80 attributes were confirmed by the SMEs as adequate to prove a fraudulent exports. These 80 attributes went through the changes – using Gradient Boosting Machine (GBM) author identified 18 attributes to explain 80% variability, and 18 attributes were relativized to create 18 indices. These 18 indices reflect the participation of the attribute in which anomalies are sought. To detect anomalies, two methods are evaluated: AutoEncoder and PCA. AutoEncoder is found to be 20 times faster than PCA, AutoEncoder finds the subtle anomalies which PCA fails to detect and can detect the anomalies even with higher latent dimensions that linear PCA cannot. PCA requires more computation than AutoEncoder. Mean Square Error method is used to measure how distant the predictions were from the real data. Author has acknowledged the difficulty in evaluating the results against the business objectives to be achieved. This work has identified the suspected cases of fraudulent exports using unsupervised deep learning that looks promising; however, it needs further thorough assessment by experts. Depending on the assessment results by experts, the model may need to adjust the number of hidden layers and neurons to have better results.

- **Graph CNN**

A novel approach, based on the graph CNN was presented by (Alarab et al. 2020) to predict illicit transactions in the bitcoin transaction graph. The proposed method has used GCN along with multi-layer perceptron, which has given better results than only GCN as used in the original research paper. The features derived from GCN and the latent representation of a linear layer boosts the performance of the model. The proposed approach is evaluated using elliptic data, a publicly available dataset that belongs to real Bitcoins transactions, which is represented as a directed graph network of transactions that are nodes, the directed edges between the transactions represent the payment flow from source to destination. The dataset is labelled as licit and illicit. The graph network containing 203,769 node transactions and 234,355 edges representing the payments flow between nodes. 2% of the data set are labelled as illicit, while 21% are labelled as licit

transactions. Accuracy of this model is 0.974 while the accuracy of previous models compared is 0.961; however, the difference is statistically insignificant. Real time of the transaction is not used as a feature in the proposed model, which is identified as an additional feature for the model to try out in the future.

- **Multi-Layer Perceptron**

A multi-classification method for detecting suspicious behaviour and categorizing the money laundering related crime using SVM, Liner Regression (LR) and Multi-Layer Perceptron was proposed by ("16th Web Information Systems and Applications Conference, WISA 2019" 2019). Two models are produced as a part of it. Model 1 trained using SVM, LR and MLP (with ReLU activation function) methods by giving inputs as financial transactions and profile data, that gives the suspicious transaction dataset as output, and Model 2 is trained using the same methods, by giving the input dataset as fraudulent transactions and suspicious transactions to determine the suspicious transactions along with the crime category. Author used the synthetic data for training and testing. The model precision rates are – 88.13, 87.53 and 90.42 for SVM, LR and MLP, respectively.

2.3.3 Interpretability Analysis

Advancement in computing and storage infrastructure, continuous generation of large volume of data and a need to utilize this data to derive actionable insights has escalated the research in machine learning methodology over the last decade. The accurate prediction has remained the primary goal during the model development and ended up having many complex models or black box models, such as deep learning models with high accuracy. These complex models are opaque, whose predictions are difficult to understand for humans. The lack of transparency, interpretability and explainability of the deep learning/machine learning models is making the adoption difficult for many business domains (medical, banking, crime, law) where the predictions or recommendations made by black boxes are expected to be used to make critical decisions that involves human life. Hence, interpretability or explainability is essential to advance the adoption of machine learning and deep learning. Not every domain requires the explainability, for example: the weather forecast does not need explanation as long as the

model predicts highly accurate results. Similarly, advertisements displayed on social media based on the browsing habit do not need the explanation of how the ads are chosen.

XAI has become a prerequisite for building the trust, and to drive adoption of AI systems in high stake domains such as finance crime, credit risks, healthcare which requires reliability, safety and fairness (Mashrur et al. 2020), (Königstorfer and Thalmann 2020) (PricewaterhouseCoopers 2017). The regulations like General Data Protection Regulation (GDPR) (Union 2016) particularly 'Records of Processing Activities', and 'Right to be informed'; and the California Consumer Private Act (CCPA) (State of California 2018) have also imposed the interpretability and explainability mandates for most of the AI/ML solutions in regulatory compliance domain.

Many methods are found in literature with reasonable accuracy such as, detecting the money laundering patterns (Savage et al. 2017a), suspicious transactions (Jullum et al. 2020), (Zhang and Trubey 2019), (Magomedov et al. 2018), anomalies (Guevara et al. 2020a), (Paula et al. 2017b); or for performing the investigative tasks using link analysis (Han et al. 2015a) or graph mining (Kanezashi et al. 2019).. For AML domain, where the suspicious transactions are required to be reported to regulatory by preparing the SMRs, it is critical to include adequate evidences that justifies the suspiciousness. Since most of the machine learning models that give high accuracy are black box in nature, it fails to create the trust in using AI/ML enabled system.

To assess the interpretability of the machine learning models applied in AML domain, we conducted a study of 77 research articles (identified in a "Methods in AML" category as per **Table 2.2**) that presented a unique model for detection of money laundering. We choose 43 papers that described the machine learning methods, and excluded the papers that described approach, system architecture or papers from other domains. As shown in the **Table 2.5**, methods are classified using supervised, semi-supervised, unsupervised and reinforcement learning category. The *problem* column indicates the key problem addressed by *methods* using *data* in the paper mentioned in *Reference* column. The meaning of *interpretability* in column 'Is Method Interpretable' is, the ability of the method to present indicators that can help humans to understand the functioning of the method to know how decisions are made by the method.

Table 2.5 – Interpretability of methods used for AML domain

No.	Category	Problem	Method	Is Method Interpretable	Data	Year	Reference
1.	Supervised learning	Detect suspicious transaction using active learning	Isolation and RF	Yes	Synthetic	2022	(Labanca et al. 2022)
2.	Supervised learning	Watch-list filtering automation to prevent false positives	SVM, Naïve Bayes, Decision Tree - individually	Yes	Synthetic	2021	(Alkhalili et al. 2021)
3.	Supervised learning	Detect suspicious transactions	Logistic regression, Decision Tree, Gradient Boosting Trees	Yes	Real	2021	(Baesens et al. 2021a)
4.	Supervised learning	Detect suspicious transactions	RF, Decision Tree, SVM, Logistic Regression, Naïve Bayes	No*	Synthetic	2021	(Tundis et al. 2021)
5.	Supervised learning	Identification of illicit Bitcoin transactions	Ensemble Method using RF, Extra Trees, and Bagging Classifier.	No	Elliptic	2020	("ICMLT 2020 - Proceedings of 2020 5th International Conference on Machine Learning Technologies" 2020)

No.	Category	Problem	Method	Is Method Interpretable	Data	Year	Reference
6.	Supervised learning	Predict illicit transactions in bitcoin transaction graph	Graph Convolutional Network and Multi-Layer Perceptron together.	No	Elliptic	2020	("ICMLT 2020 - Proceedings of 2020 5th International Conference on Machine Learning Technologies" 2020)
7.	Supervised learning	Detecting potentially illicit behaviour	Logistic Regression, *XGBoost trees Catboost methods*	No*	Real	2020	(Tertychnyi et al. 2020)
8.	Supervised learning	Detect illicit accounts involved in money laundering over ethereum blockchain	XGBoost	No	Public Dataset	2020	(Farrugia et al. 2020)
9.	Supervised learning	Examine AI methodologies against money laundering crimes	Logistic regression, decision tree, *RF, XGBoost, and AutoEncoder, DNN*	No*	Real	2020	(Garcia-Bedoya et al. 2020)
10.	Supervised learning	Detect suspicious money laundering transactions	XGBoost	No	Real	2020	(Jullum et al. 2020)
11.	Supervised learning	Detect suspicious money laundering transactions	Naïve Bayes Classifier	Yes	Synthetic	2020	(Kumar et al. 2020)

No.	Category	Problem	Method	Is Method Interpretable	Data	Year	Reference
12.	Supervised learning	Identification of suspicious transactions and categorizing the type of finance crime	SVM, Liner Regression, *Multi-layer Perceptron*	No*	Synthetic	2019	("16th Web Information Systems and Applications Conference, WISA 2019" 2019)
13.	Supervised learning	Suspicious money laundering transaction detection	Bayes Logistic Regression, Decision Tree, RF, SVM, and *ANN*	No*	Real	2019	(Zhang and Trubey 2019)
14.	Supervised learning	Suspicious money laundering transaction detection	RF; MinMaxScaler method	Yes	Synthetic	2018	(Magomedov et al. 2018)
15.	Supervised learning	Proposes a technique for generating variants of typologies	Graph Learning, BigData	No	Unclear	2018	("Proceedings - 2018 IEEE 6th International Conference on Future Internet of Things and Cloud Workshops, W-FiCloud 2018" 2018)
16.	Supervised learning	Detect money laundering criminals – actual court case	Logistic Regression, Decision Tree, RF and *Neural Networks*	No*	Real	2018	(Badal-Valero et al. 2018a)
17.	Supervised learning	Detect money launders laundering groups	SVM and *RF*	No*	Real	2017	(Savage et al. 2017a)

No.	Category	Problem	Method	Is Method Interpretable	Data	Year	Reference
18.	Supervised learning	Improve the suspicious transaction signalling process by client profiling	PART algorithm that implements Decision tree, best leaf technique	Yes	Real	2016	(Alexandre and Balsa 2016)
19.	Supervised learning	Suspicious money laundering transaction detection	Affiliation Mapping Calculation and Sequential Mining (AMC-SM)	Unclear	Unclear	2016	(Jayasree and Siva Balan 2016)
20.	Supervised learning	Identification of suspicious money laundering accounts	Probabilistic Relational Model using the Audit Sequential Pattern (PRM-ASP) Mining	Unclear	Real	2015	(Jayasree and Siva Balan 2015)
21.	Supervised learning	Detection of fraud chains in Mobile Money Transfer systems.	Predictive Security Analysis at Runtime (PSA@R)	Unclear	Synthetic	2014	(Zhdanova et al. 2014)
22.	Supervised learning	Identification of suspicious money laundering transaction	DBSCAN Clustering algorithm, Link analysis	Yes	Real	2014	(Yang et al. 2014)
23.	Supervised learning	Identification of suspicious financial transactions	Clustering, Dynamic Bayesian Networks, Anomaly Detection	Yes	Real	2011	(Raza and Haider 2011)
24.	Supervised learning	Suspicious money laundering transaction detection	Graph learning method	No	Synthetic	2011	(Michalak and Korczak 2011)
25.	Supervised learning	Identify suspicious money laundering activities	Decision Tree (BIRCH and k-means)	Yes	Unclear	2011	(Liu et al. 2011)

No.	Category	Problem	Method	Is Method Interpretable	Data	Year	Reference
26.	Supervised learning	Suspicious money laundering transaction detection	Clustering (K-means), *Neural Network (back-propagation)*	No*	Real	2010	("Digital Forensics and Cyber Crime - First International ICST Conference, ICDF2C 2009, Revised Selected Papers" 2010)
27.	Supervised learning	Identify the most critical classifiers for decision tree used in investigation of money laundering	Decision Tree	Yes	Real	2010	(George and Kavakli 2010)
28.	Supervised learning	Identify money laundering cases within investment activities	Clustering, *Neural Network*, Genetics Algorithm	No*	Real	2010	(Le Khac and Kechadi 2010)
29.	Supervised learning	Identification of suspicious money laundering transaction	Decision Tree	Yes	Real	2008	(Zhang and Zhao 2008)
30.	Supervised learning	Detection of hidden money laundering behaviour	*Multi-agent Neural Network*, *Text Mining*, Genetic Algorithms, Velocity Analysis and Case-based reasoning	No*	Real	2007	(Yang et al. 2007)
31.	Supervised learning	Facilitate investigation of money laundering crime	Link discovery based on correlation analysis	Yes	Real	2003	(Zhang et al. 2003)

No.	Category	Problem	Method	Is Method Interpretable	Data	Year	Reference
32.	Unsupervised learning	Suspicious money laundering transaction detection	Core Decision Tree and Clustering Algorithm	No*	Real	2021	(Rocha-Salazar et al. 2021)
33.	Unsupervised learning	Identify fraudulent financial transactions	Outlier Detection Methods, Visual Analytics Method	Yes	Real	2020	(Torres and Ladeira 2020)
34.	Unsupervised learning	Identify anomalies in a set of transactions of a non-banking correspondent	Isolation Forest, One Class SVM	Yes	Real	2020	(Guevara et al. 2020a)

No.	Category	Problem	Method	Is Method Interpretable	Data	Year	Reference
35.	Unsupervised learning	Detect hidden networks of money launderers	Isolation Forest – One Class SVM	Yes	Synthetic	2020	("Proceedings of the 13th IADIS International Conference ICT, Society and Human Beings 2020, ICT 2020 and Proceedings of the 6th IADIS International Conference Connected Smart Cities 2020, CSC 2020 and Proceedings of the 17th IADIS International Conference Web Based Communities and Social Media 2020, WBC 2020 - Part of the 14th Multi Conference on Computer Science and Information Systems, MCCSIS 2020" 2020)

No.	Category	Problem	Method	Is Method Interpretable	Data	Year	Reference
36.	Unsupervised learning	Identify suspicious transaction in crypto currency	Expectancy Maximization (for clustering datasets), RF (for anomaly detection)	No	Elliptic	2019	(Baek et al. 2019)
37.	Unsupervised learning	Detect anomalies considering the regular transaction patterns	AutoEncoder and PCA examined separately	No	Real	2017	(Paula et al. 2017b)
38.	Unsupervised learning	Identify suspicious money laundering transactions	*Ensemble* of algorithms (Isolation Forests, One Class SVM, *Gaussian Mixture Models*, EM)	No*	Real and Synthetic	2017	(Camino et al. 2017)
39.	Unsupervised learning	Detection of money launderers gangs	Community detection using temporal-directed Louvain algorithm	No	Real	2017	(Li et al. 2017)
40.	Unsupervised learning	Detection of money laundering transactions. Visualization and analysis system for Police Analysts.	Apriori, PrefixSpan, FP-growth and Eclat algorithms	Yes	Real	2016	(Drezewski et al. 2016)
41.	Unsupervised learning	Suspicious money laundering transaction detection	Clustering CLOPE algorithm	Yes	Real	2012	(Cao and Do 2012)
42.	Unsupervised learning	Identify suspicious sequences of transaction level for financial institutions	Scan Statistics	Yes	Real	2010	(Liu and Zhang 2010)

No.	Category	Problem	Method	Is Method Interpretable	Data	Year	Reference
43.	Unsupervised learning	Detect suspicious money laundering transactions	Semantic Core Tree	Yes	Synthetic	2006	(Chen et al. 2006)
44.	Semi-supervised learning	Investigation Support System for AML	Multi-channel Convolutional Network based on NLP	No	Synthetic	2015	(Han et al. 2015a)
45.	Reinforcement learning	Incremental graph pattern matching algorithm to deal with time-evolving graph data for detecting money laundering	Incremental graph pattern matching (IGPM) algorithm and partial execution manager (PEM) to re-compute the updated subgraphs	No	Temporal graph data	2019	(Kanezashi et al. 2019)
46.	Supervised and unsupervised learning	Detect trade-based money laundering	Supervised – Bayesian network, Cost sensitive learning Unsupervised – clustering, regression	Yes	Real	2019	(Chao et al. 2019)

* It indicates that the methods used by authors of respective paper includes interpretable and non-interpretable methods.

Figure 2.6 to **Figure 2.9** show the analysis based on the 43 papers filtered during review that provides the solutions for money laundering pattern detection, suspicious transaction detection, money launderers' group detection and link analysis, using machine learning.

Figure 2.6 shows that the highest number of papers are published in year 2020 in AML domain that indicates the escalation in attention the subject is getting to leverage machine learning technology to help solve one of the highly complex problem of industry.

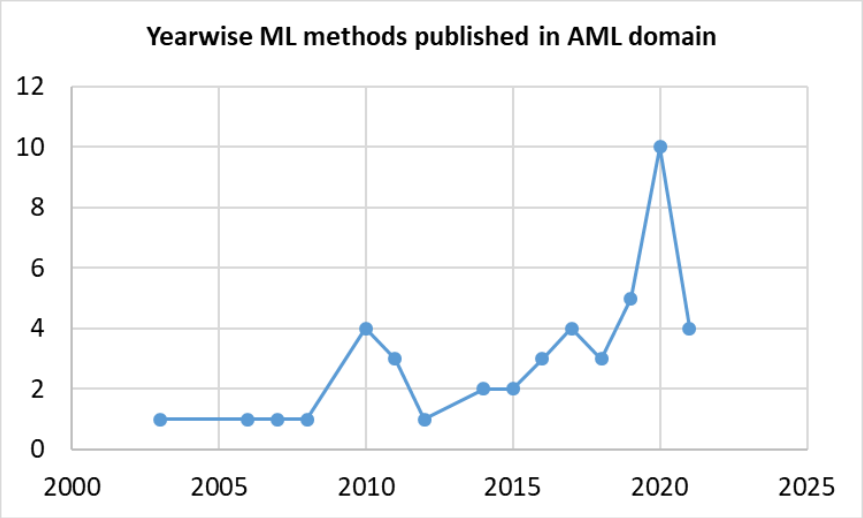


Figure 2.6 – Year wise machine learning methods published in AML domain

Figure 2.7 shows that 51% researchers have used non-interpretable methods.

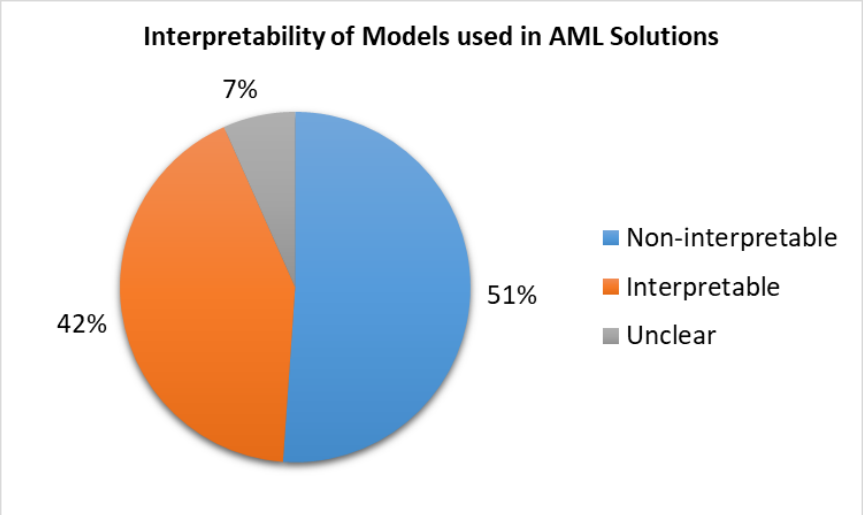


Figure 2.7 – Interpretability of machine learning methods used in AML solutions

Figure 2.8 shows 65% methods are based on supervised learning while only 2% have used reinforcement learning models.

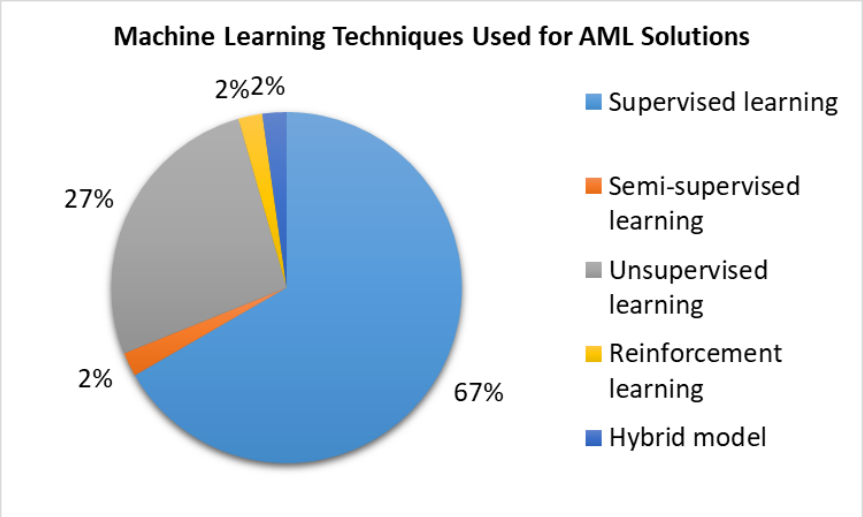


Figure 2.8 – Category of machine learning techniques used for AML solutions

Figure 2.9 shows 58% methods are either trained or evaluated on a small set of real data older than 2 years. 22% researches are done based on synthetic data. It is observed that the real data is obtained from following organizations – financial institutions, financial intelligence units, police department, and organizations involved in exports.

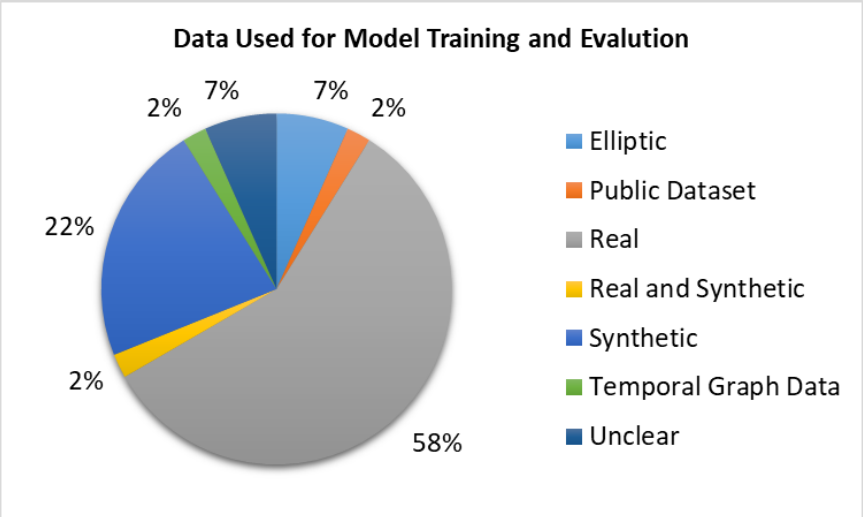


Figure 2.9 – Type of data used for model training and evaluation

Given the steps involved in AML, efforts involved by multiple organizations together, it becomes more critical to identify the suspicious money laundering transactions, patterns, groups as accurately as possible along with adequate explanation for the decisions made by the machine learning models. As the model accuracy increases the interpretability goes down, and in order to know how models are making decisions it needs to be interpretable. So far, the research focus has been more on increasing the model accuracy and rather than ensuring the model interpretability. This has triggered another research area that is *XAI*.

2.3.4 Explainable AI Techniques

XAI is a class of system that provides visibility into how an AI system makes the decisions, predictions and executes its actions (Rai 2020). Inter-disciplinary research area, XAI has attracted great attention from the researchers around the world. XAI techniques are applied in other related domains such as, in bitcoin domain (Shah et al. 2021) presented a study of applying XAI techniques to visualize and understand the results obtained from unsupervised learning, (Westerski et al. 2021) presented an explainable anomaly detection technique for procurement fraud detections; however, no research found in the literature that describes the application of XAI techniques for money laundering domain. This is a clear gap and potential for future research.

The study conducted by (Cirqueira et al. 2020) enforces the need to understand the operational environment of stakeholders who are going to use the explanations generated by XAI methods, to support their decision-making process. Author proposes a scenario-based requirement elicitations method for developing user-centric explanations using XAI methods for fraud detections. Knowing the lack of interpretability and explainability of the black box models, researchers have already made exponential progress in coming up with different post-hoc approaches over last 5 years. There are several review papers published on XAI techniques that describes concepts, taxonomy, methods, and challenges; the most recent comprehensive review papers on XAI subject are listed in **Table 2.6**.

There are several reasons why the explanation is required for the decisions made by black box models based on the subject of research domain, for example – regulatory application domain wants the explanation to *justify* the decisions, security application domain wants

the explanation to *control* the underlying data and IP that may be uncovered, some domains would need the explanation to *improve* the decisions, other domains would need the explanation to discover the knowledge (Adadi and Berrada 2018). Further authors discussed the topic using 5 W's and How (What, Who, When, Why, Where, and How) to bring out the existing literature.

Table 2.6 – XAI review papers

No.	XAI review paper title	Reference
1	Peeking Inside the Black-Box: A Survey on XAI	(Adadi and Berrada 2018)
2	XAI: A survey from trust and genuine need of explainability perspective	(Dosilovic et al. 2018)
3	XAI: History, Research Areas, Approaches and Challenges	(Xu et al. 2019)
4	XAI: Concepts, taxonomies, opportunities and challenges toward responsible AI	(Barredo Arrieta et al. 2020)
5	Extracting Relational Explanations from Deep Neural Networks: A Survey from a Neural-Symbolic Perspective	(Townsend et al. 2020)
6	Explaining the black-box model: A survey of local interpretation methods for deep neural networks	(Liang et al. 2021)
7	A Survey of Data-driven and Knowledge-aware XAI	(Li 2020)
8	A Survey of Contrastive and Counterfactual Explanation Generation Methods for Explainable Artificial Intelligence	(Stepin et al. 2021)
9	XAI: A review of machine learning interpretability methods	(Linardatos et al. 2021)
10	Explainable Reinforcement Learning: A Survey	(Puiutta and Veith 2020)

The survey, conducted by (Dosilovic et al. 2018), is emphasizing the literature based on the importance of the trust in the system and a genuine need of explainability, Transparency and Interpretability for the AI applications in judiciary, governmental, finance and autonomous transport.

XAI work is categorized into two groups by (Xu et al. 2019) – (i) *transparency design* that focuses on the internal functioning of the model from developer's point of view and (ii) *post-hoc design,* which explains why a result is inferred from the user's point of view.

The study conducted by (Barredo Arrieta et al. 2020) has explained the underlying concepts of model Explainability, XAI taxonomies, opportunities, challenges and way forward to Responsible AI. XAI purpose and concepts are discussed using *what*, *why*, *what for*, and *how* aspects. The review of XAI is presented using two categories – (i) *potentially interpretable models* - the machine learning models having some level of transparency to make it interpretable (such as linear/logistic regression, decision tree, K-nearest neighbours, rule based learning, general additive models and Bayesian models) and (ii) *post-hoc techniques* (model agnostics, post-hoc techniques for shallow models, Explainability in deep learning, and alternate taxonomy for post-hoc techniques for deep learning models) that helps to make the machine learning models interpretable; resulting in a global taxonomy. Further, the author leads the discussion towards *Responsible AI*, by defining the AI principles (fairness, transparency, and privacy) to follow when implementing AI models. It is noted that implementing XAI has some implications, especially on compromising the privacy involved by the data and business rules which are confidential in nature. Author's view is, the model interpretability should be addressed together with requirements and constraints of data privacy, model confidentiality, fairness and accountability.

Researchers from neural-symbolic integration (NSI) have tried to extract the relational knowledge from neural networks by proposing several methods but have remained a challenge to extract distinct and salient features from the input and hidden layers. Further, the methods of identifying the relationship between the features have emerged. Weber et al. (Townsend et al. 2020) presented a review of the old and new methods of extraction without relational explanation. The review presents 29 knowledge extraction methods by grouping them into 3 categories – (i) work that extracts quantized features only, (ii) work that extracts quantized feature along with relationship and (iii) embedding methods with associated extraction methods.

Due to the surge in performance of deep neural network models but lagging on explainability, has driven many researchers to come up with different machine learning interpretable methods. Significant efforts have gone into producing different interpretable methods, out of which the local interpretable methods are standing out from clear expression of feature and low on computation complexity point of view (Liang et al. 2021). The local interpretable methods for DNN published between 2012 and 2020 are categorized into two groups (Liang et al. 2021)as (i) model-driven – describes 20 methods and (ii) data-driven – describes 27 methods.

Weber et al. (Li 2020) reviewed a literature of XAI published over the last decade and presented using three phases of XAI lifecycle – methodology, evaluation and application. The methodology is presented using two categories – (i) *knowledge driven XAI* which requires external knowledge to produce the explanation, and (ii) *data driven XAI* methods which are reviewed from global, local and instance-specific methods point of view. The XAI evaluation metrics are classified into two categories – computational metrics (usually standardized and automated by AI experts) and cognitive metrics (usually collected by user-study). Further author has reviewed the XAI application for self-driving and finance AI.

There are multiple ways to look at the explanation required for the decisions made by black-box models. One approach is *evidence-based explanation* that describes the details of parameters that contributed to make the decision making. Another approach is *Contrastive and Counterfactual (C&C) explanations* that justify why the output of the algorithms is not any different from what it is and how it could be changed, respectively. (Stepin et al. 2021) examined the theoretical foundation for C&C explanations and reported the state-of-the-art computation frameworks for generating the C&C explanations and identified the shortcomings which can be a topic of future research.

Weber et al. (Linardatos et al. 2021) presented a comprehensive review of machine learning interpretability methods using 4 categories – (i) Methods for explaining black box models (17 methods focused on deep learning method interpretations, 16 methods that can explain any black-box model), (ii) methods for creating white-box models (5 methods), (iii) methods that promote fairness and restrict discrimination, (iv) methods

that analyses the sensitivity of model predictions (28 methods). Most interpretability methods are focused on deep learning, largely ruled by neural networks, and are experimented with image classification explanation. LIME (Ribeiro et al. 2016) and SHAP (Lundberg and Lee 2017) are the highly referred and experimented methods for producing explanations on black box models that use any type of data, followed by Partial Dependence Plots (PDPs).

Weber et al. (Puiutta and Veith 2020) conducted a XAI literature review by keeping the reinforcement learning at centre, because most of the XAI literature reviews are around supervised learning. In reinforcement learning, the models autonomously learn and make the decisions, which is equally important to know the reasons or explanations behind the decisions.

Knowing the need and importance of having the answers for following questions to AI models – why did you do that, why not something else, when do you succeed, when do you fail, when can I trust you, how do correct an error; USA Defence Advanced Research Project Agency (Gunning and Aha 2019) initiated a 4 year long XAI program to work on the same.

While there is mammoth progress on XAI method, (Rudin 2019) argues that the development in XAI is a patchwork to existing machine learning models and a major effort of research community is going in the wrong direction. Author firmly recommends to stop using the black box machine learning models along with XAI techniques for high stake decisions, instead use the interpretable models. Author believes that the research effort should be directed towards developing new interpretable models that can give the required accuracy.

2.4 Discussion and Recommendations

This section describes the strengths and limitations of the models used in research so far, current challenges and constraints for researchers from data perspective, barriers for adoption of machine learning or deep learning in the AML domain and the future research directions.

2.4.1 Strengths and Limitations of Applied Models

Money laundering detection solutions are categorized in two groups. First category is to *identify the suspicious transactions*, example - (Alarab et al. 2020; Paula et al. 2017a) has presented the AutoEncoder and Graph CNN deep learning methods respectively to identify suspicious transactions; and second category is to help investigate the identified suspicious transactions or alerts identified by rule-based systems, which is commonly called as *decision support systems*, example - (Han et al. 2015b; Mark Weber 2018) has presented a multi-channel CNN using NLP and scalable GCN method as a decision support system for investigating the alerts, respectively.

Presently, many financial institutions use rule-based systems for detecting the transactions (beyond certain threshold values, international fund transfers and suspicious transactions of any nature) that are required to be reported to regulatory to be compliant with AML/CTF policies. While rule-based-system is crucial, but it generates many false positive alerts for suspicious transaction which requires huge manual effort to triage the same (example - in a global bank approximately 10,000 analysts investigates over 400,000 alerts per week (Mark Weber 2018)). Few technology savvy financial institutions have started leveraging AI/ML technology to automate such activities to improve the accuracy of detection of suspicious activities (Finance 2018) but it remains a long way for adoption by many because of lack of explainability of the decisions made by AI/ML models.

Over the last decades, several papers have been published with different machine learning techniques summarized in many comprehensive literature review papers such as (Irofti et al. 2021; Milind Tiwari 2020; Labib et al. 2020; Alsuwailem and Saudagar 2020); however, by looking at the penalties issued by authorities for financial institutions in single year 2019 (Monroe 2020), it is evident that the published methods are either not useful or not used.

Despite the technological advancements from statistical methods, machine learning to deep learning, severity of 'suspicious money laundering transactions detection' issue has remained more or less constant since last two decades (example weber et el. (Tang and Yin 2005) presented a SVM based method in 2005, (Jullum et al. 2020) presented

machine learning based technique in year 2020). Many research articles are published presenting the application of these methods to detect suspicious transactions and claimed the improvements in double digits; however, rare evidences on practical use of these methods by relevant entities such as financial institutions and the actual performance in terms of false positive alerts reduction or operational cost reduction.

Graph Neural Networks is observed as one of the most commonly used method for detection of suspicious transaction and money laundering networks. Along with the potential of using graph neural networks, there are several challenges (Kurshan et al. 2020) such as complexities due to high-speed transaction systems, real-time systems, multi-channel updates, data size, data speed, data variety, and several business applications involved; which makes the practical implementation harder. Hence, it is essential to consider the right infrastructure, appropriate tools and considerations for these challenges to improve the performance of future graph-based solutions.

2.4.2 Current Research Challenges and Constraints

There are several challenges and data constraints for conducting research on the detection of money laundering.

2.4.2.1 Current Research Challenges

Challenges exists in money laundering domain for many years primarily due to the nature of the domain being highly sensitive and confidential in nature. Few key current challenges are described as below:

- **Changing canvas** - The technological advancements have made the banking easy for customers, but also presented new avenues for fraudsters. Digital banking has changed the operating canvas and triggered an increase in volume and frequency of transactions, enabled multiple customer channels, fastened real-time transaction settlement. With digital banking, the fraud patterns are also changing; and regulatory landscape is changing due to expansion of banking across the countries and complexities are increasing due to new or change in regulatory policies.
- **Multi-step classification** – The inherent nature of 'suspicious money laundering transaction detection' is a multi-step classification problem, where the decision of

transaction being suspiciousness is taken at multiple levels that includes transaction level, alert level, case level, SMR level and finally by the law & enforcement agency level. Few decisions are taken by software and few requires human intervention for investigation. Only a degree at which the transaction is suspicious can be improved in a system, because whether the transaction is indeed money laundering or legitimate is always decided by a well-established and responsible government agency at the end.

- **Slow adoption of AI technology** – Though many complex models such as deep neural networks have given promising results in terms of accuracy and performance in other domains, but lacks on the transparency and interpretability of models which creates a barrier for AI adoption in AML domain.

- **AML domain being confidential in nature** – Many researched methodologies for detecting suspicious money laundering transactions are patented and hence they are not published for obvious reasons (Korauš et al. 2019). This also creates a gap/disconnect between what is indeed applied in practice vs. what techniques/methods available, to solve the problem.

2.4.2.2 Data Constraints

For conducting any research, data is of utmost importance and AML domain from a finance industry is one of the highly sensitive domain, and that brings many constraints for researchers from data perspective as described below:

- **Real data** – Banks and financial institutions are the custodians of the customer's data that includes customer's personally identifiable information, the products they hold and transactions data associated. There are various regulations in place such as Global Data Protection Regulation (GDPR) (Union 2016), California Consumer Private Act of 2018 (State of California 2018) etc. The customer data (limited to own data) cannot be shared with anybody other than the customer themselves. From research of "suspicious transaction detection" point of view, customer, products and transactions are the key source of data. The researches where real data is used, mostly it is older than two years and that too a fraction of the data that any financial institution possesses.

- **Training data** – The suspicious transactions data and suspicious transaction alerts data is considered as highly sensitive data, since this is used by banks to prepare SMR for submission to regulatory. Based on the investigation outcome of the SMR by regulatory, it results in either SMR close or criminal investigation against the account holder. Banks are obliged to keep the SMRs confidential from even the customer who owns the data (Office of Parliamentary Counsel 2019). From a research point of view, suspicious transaction data and suspicious transaction alerts data are key label data to train the supervised machine learning model to detect suspicious money laundering transaction which are again not easily available for researchers unless the research is collaboratively conducted with financial institutions.

- **Authenticity of training data** – The transactions involved in actual money laundering is rarely known to the financial institutions, because it remains confidential information with law enforcement department and regulatory (Office of Parliamentary Counsel 2019). This is an accurate label data to train the supervised machine learning models to detect suspicious money laundering transaction, but is nearly impossible to get it for research purpose unless the research is collaboratively conducted with FIUs.

- **Real data speed and volume** – Every day, the bank generates millions of transactions of different types through various banking systems and multiple channels. This is a large volume of data, generated at large velocity and of large variety. Many banks are struggling to catch up with the speed at which the transaction data is getting generated and consolidating the same together to get a view for analysing it to gain the insights or apply the advanced analytics techniques to identify the money laundering and financial frauds.

- **Class imbalance, overlap and mislabelling** – In the daily set of transaction data at financial institutions, the suspicious transactions are way too less (a fraction of transactions) than the legitimate transactions that creates a data imbalance while training the machine learning models. Money launderers smartly split the transactions in a way which appears similar to legitimate transaction that creates

the overlapping transactions and makes it difficult to identity true suspicious transactions. Unavailability of the true money laundering case data, forces the researchers to rely upon the existing suspicious transaction data which may results in mislabelling of the data used for machine learning models (Sudjianto et al. 2010). A study conducted by (Tertychnyi et al. 2020) has presented a two-layered approach to address the class-imbalance and scalability issue for machine learning.

2.4.3 Barriers for Adoption of AI/ML in AML Domain

Financial institutions believe in the potential of AI technology that it can help in several business scenarios and use cases (Königstorfer and Thalmann 2020). Some organizations have already started taking baby steps by developing small prototypes using AI/ML (Finance 2018). However, the overall adoption of AI is slow primarily due to following reasons:

- **Explainability** - the models that give high prediction accuracy are usually non-interpretable in nature that makes it difficult to understand the reasoning behind the decisions made (Mashrur et al. 2020) by machine learning models.
- **Trust** – humans trust the system if they know exactly how it functions. The black-box models which gives high accuracy but lacks on interpretability of the model, does not inspire the confidence and trust in the system (PricewaterhouseCoopers 2017).
- **Privacy** – the data used by machine learning algorithms includes the customer, accounts and transaction data. There are concerns around the usage of this information and revealing the confidential information while bringing out the explanations of the decisions (Board 2017) made by machine learning or deep learning model, by using XAI techniques.
- **AI ethics** – there are concerns regarding the AI agents making the biased decisions or discriminations while interpreting and making decisions (Board 2017).
- **Law and enforcement** – the courts across many developed countries are struggling to develop convincing and clear-enough guidelines for directing

legislative and administration considerations for adoption of AI (Rademacher 2019).

- **Data quality** – data quality has been a critical issue. As per the study performed by (Gupta et al. 2022), identifies the issues such as incomplete and inaccurate data, unavailability of the data, the time required for processing the alerts and cases that goes from 15 days to 10 months causing a different insights into data.

2.4.4 Future Research Directions

Machine learning and deep learning technology is used by select few AML software product vendors (such as SAS, Actimize, Oracle, LexisNexis, Fiserv.) for suspicious transaction detection. Most financial institutions heavily depend on third party products for detecting money laundering suspicious activities. Different statistical methods, data mining and machine learning techniques are used by these AML products. Most AML products are trained on synthetic data / old data and once the product is live in financial institutions it is fine-tuned based on the real data. The third party products are anyways IP protected but on top of it, even the decisions made by the products are not transparent and explainable to AML analysts to inspire the trust and confidence in system. Hence, the research on AML domain using machine learning or deep learning should be driven by 'Responsible AI' principles to ensure the technology is applied in transparent way, yet safeguarding the interest of each participant involved in the ecosystem.

While there is a continuous research on finding out the best possible solution to detect and report the money laundering transactions, by researchers, commercial product companies and financial institutions; following are the future research directions identified after studying the literature:

- **Data pre-processing** – AML domain uses the data generated by internal banking systems, data from authorised third party such as World-Check, PEP screening, Sanctions screening, and social network data. Most of the published researches have used a limited set of data (limited to transactions) and ignored the data pre-processing part (Chen et al. 2018c) and (Lokanan 2019). It is important to get the complete domain context, data sources, data types, and leverage the same to detect money laundering.

- **Evaluating the model against large data** – Most of the methods found in the literature have been evaluated using a small set of data on limited computing, compared to the data that gets generated practically on day-to-day basis in financial institutions. Hence, it is important to consider the large data scenarios to evaluate the models.

- **Graph mining and Social network analysis** – These techniques are an apt solution for finding the patterns, groups and perform the link analysis ("Proceedings - 2018 IEEE 6th International Conference on Future Internet of Things and Cloud Workshops, W-FiCloud 2018" 2018), which is helpful for detection of money laundering, launderer gangs (Savage et al. 2017a) and identify the relationships to find more leads in the money laundering networks. There are challenges in graph mining from large data processing point of view (Kurshan et al. 2020), and this is one of the potential area where research should be continued that can help improve decision support system for AML investigators.

- **Applying XAI techniques** – It is a need of AML domain to have an Explanation of the decisions made by non-interpretable models (Mashrur et al. 2020), (Board 2017). There are several XAI techniques researched so far, for example few of them are - (Gunning and Aha 2019; Lundberg and Lee 2017; Rai 2020; Ribeiro et al. 2016), however, to the best of our knowledge none of it is applied in AML domain yet. Hence, application of XAI techniques can be further researched on the non-interpretable models that gives a good accuracy of detection.

- **Ensemble learning** – Considering the number of financial systems, type of available data, data sources, volume of data; one specific model cannot be a solution, instead ensemble learning must be considered to leverage the available data to bring out more confidence in suspiciousness of transaction along with evidences. It could include statistical methods, data mining, machine learning and deep learning all together. Example - (Camino et al. 2017) has applied the ensemble learning to find the suspicious transactions in distributed ledger payments.

- **Unsupervised learning** – Research efforts can be better utilised by focusing on unsupervised learning or reinforcement learning (Chen et al. 2018c), considering the fact that AML data is highly imbalanced and true labelled data unavailability (Canhoto 2020).

- **Academician and industry gap** - Many researchers attempted to solve the suspicious transaction detection issue, have used a synthetic data or a very small sample of real data compared to the actual data financial institutions generates on a day-to-day basis. Hence, the results shown by research papers can be far away from the reality if the methods are applied in practice. Part of the reason being, customer account and transaction data is highly sensitive and not always available in a required volume. It is recommended to explore wherever possible to conduct a joint research with the organisations from finance industry to benefit from each other's expertise and controlled access to real data for research.

- **Data sharing between reporting entities** – Each financial institution has the accumulated intelligence / knowledge gained about the suspicious customers, suspicious transaction patterns, and fraud patterns. If the accumulated knowledge is available to all entities with adequate legal protection to prevent misuse of the same; combating money laundering can be improved further. A study conducted by (Zand et al. 2020) presents a secure framework for AML using machine learning and secret sharing between Banks, which is one step in this direction. Another study conducted in Japan proposes a global knowledge management of suspicious money laundering transactions from financial institutions (local) to FIU (national) to FATF/ Egmont Group (international) (Yasaka 2020).

2.5 Summary

Money laundering problem exists since last five decades and the challenge of identifying suspicious money laundering transactions remains the same today as well, primarily due to following reasons - continuously changing canvas of fraud, technology and regulatory; being a multi-step classification problem, the challenges around the data for research and barriers for adoption of AI by industry. The extensive literature review suggests that there

are several machine learning methods recommended for suspicious transaction detection by researchers; however, hardly any evidence of application of the same in the industry. The possible identified reasons are – most researches are based on either sample of a very old real data or synthetic data, lack of training data (alerts, suspicious transactions) and lack of actual money laundering transaction data, lack of current real-data containing possible latest fraud patterns, and data being highly sensitive and confidential in nature. Limited research is seen on applying deep learning methods and absolutely no research in the area of applying XAI techniques for suspicious money laundering detection domain to the best of our knowledge. From the list of deep learning techniques identified from literature, out of 5 articles, 3 articles have chosen CNN variant – graph based CNN, scalable CNN and multi-channel CNN based on NLP; 4th article has chosen AutoEncoder and 5th chose MLP. This study has found that 51% of the machine learning models used in AML solutions are non-interpretable methods. Hence, along with accuracy and performance, it is important to consider model transparency, interpretability, explainability, privacy issues and define how the proposed model is going to fit into the overall AML solutions stack to put it into practice. This review has identified the key challenges for AML, barriers for adoption of AI/ML, constraints for researchers and future research directions. In a summary, apart from focusing on deep learning and XAI technology for AML domain, possible research directions for better outcome is a joint research programs between academicians and financial institutions focusing on identification of suspicious transactions; between academicians and Financial Intelligence Units (FIUs) focusing on investigation and decision support systems.

3 MATERIALS AND RESEARCH METHODOLOGY

3.1 Introduction

This chapter describes the overall research methodology and machine learning methods undertaken for the study to address research objectives. The type of data required for research, challenges in getting real data, synthetic data creation approach, design of data creation utility, and design of money laundering scenarios for training the model are discussed in detail. To address objective-1 of research, a design of CNN method for detection of suspicious money laundering transactions is presented, along with the traditional machine learning methods considered for comparing the CNN results which include RF, SVM and XGBoost. To address the objective-2 of the research, mode-agnostic XAI techniques SHAP and LIME are described which are applied on CNN classifier to explain the predictions made. Finally, the software tools used for synthetic data creation and modelling are described.

3.2 Overall Methodology

This section illustrates the design methodology of the research that includes design of method for detecting suspicious money laundering transactions using deep learning method, design of traditional machine learning methods to detect suspicious transactions, so that the deep learning method performance can be compared with machine learning methods; and design of the XAI technique to explain the decisions made by deep learning methods.

Figure 3.1 shows the design of the research development methodology followed to conduct the research. The methodology has 4 major parts – part 1 synthetic data creation, focuses on a methodology to create synthetic data containing customers, accounts and transactions in tabular format; part 2 deep learning - CNN, concentrates on developing a CNN classifier based on Conv1D layers for predicting the suspicious transactions; part 3 XAI - SHAP, centres around explaining the predictions made by CNN using SHAP XAI techniques and part 4 traditional machine learning models - focuses on comparing the

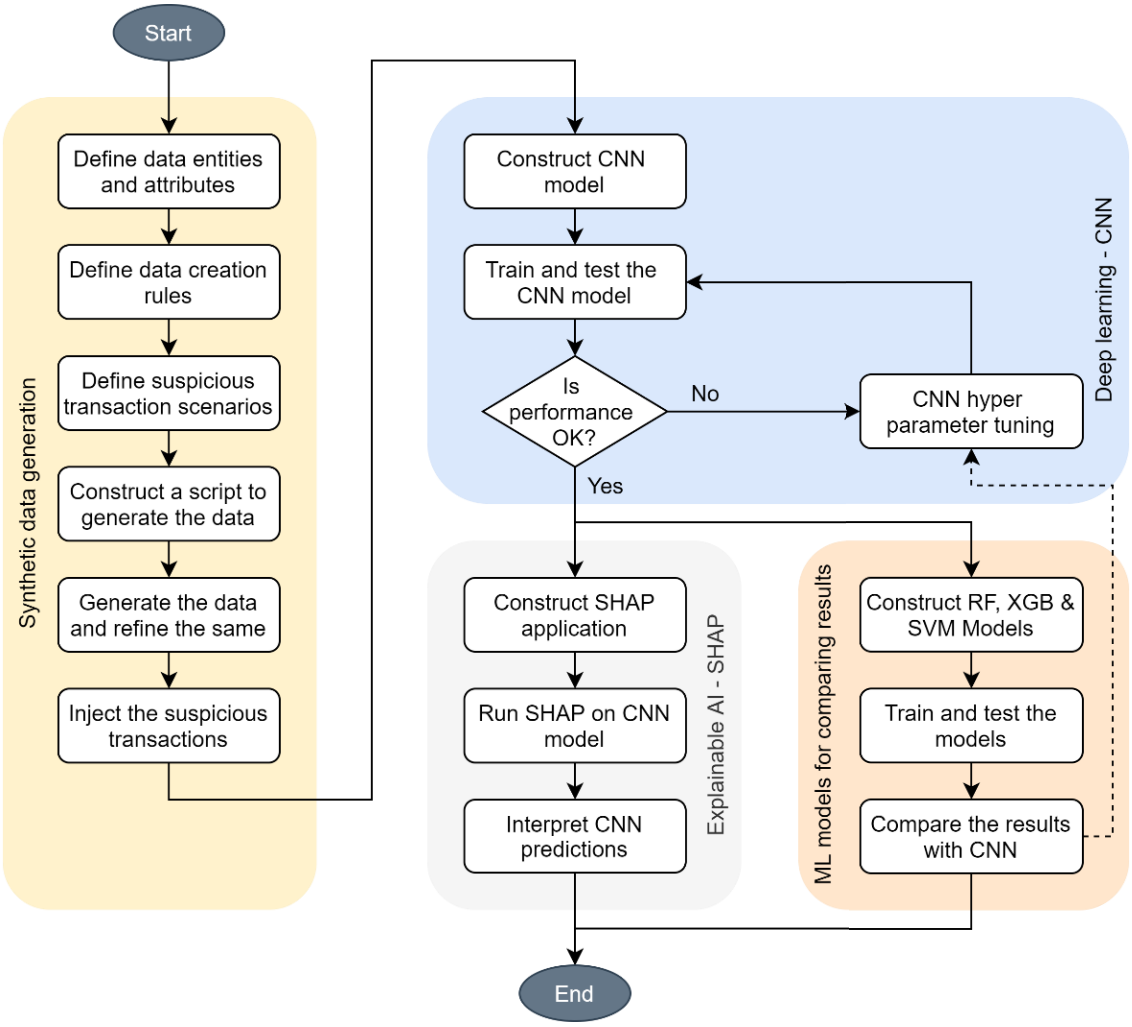


Figure 3.1 – Research development methodology

prediction performance of CNN classifier with RF, XGBoost and SVM machine learning classifiers.

3.3 Data Acquisition

Detection of suspicious money laundering transactions is carried out by reporting entities and Financial Intelligence Units (FIUs) to fight against money laundering. Reporting entities are the ones who are usually involved in providing financial services, gambling, bullion, and digital currency exchange services. Each organisation has access to different set of data.

Data used by financial institutes for money laundering detection

The following data generated by financial institutions is used to detect the suspicious money laundering activities and report the same to regulatory to ensure compliance.

- Customer – bank customer data including profile, KYC, CDD, risk profile
- Products – the products customer holds such as savings account, loan account, credit accounts, etc.
- Transactions – the financial transactions on the products that customer holds
- Transaction logs – the technical details of the transactions such as device, location, etc.

Financial institutions also use the following external datasets for risk assessment of the customer, transaction, and degree of transaction suspiciousness.

- Global sanctions lists and Politically Exposed Persons (PEPs) – there are several companies who provides the online sanctions and PEP screening services. Example – LexisNexis (LexisNexis 2022), BAE Systems (Systems 2022).
- World-Check data – it is a database of PEPs, high-risk individuals and businesses maintained by Refinitiv company and used by several organisations around the world to identify and manage the regulatory and financial risks.
- Social media data – this includes several social media websites such as Facebook, LinkedIn, and twitter, etc.

Data used by FIUs for money laundering detection

Similarly, the FIUs use the following data for investigations of the SMRs reported by reporting entities to analyse the transactions in a bigger context and detect criminal activities. Apart from the SMRs data, FIUs also have the access to the following data:

- Suspicious transactions reported by other reporting entities.
- Historical data containing investigation results.
- Financial intelligence developed overtime.
- Taxation data from Tax office.
- Criminal records data from police departments.

For the purpose of this research, suspicious money laundering transactions detection is considered for bank or financial institution.

The data produced and gathered from a customer by financial institutions is highly sensitive and confidential in nature. Similarly, the financial data produced by several banking systems to provide financial services is also a highly sensitive data from the Bank's perspective. The data required for research is identified as – customer data, products data, and transactions data. Initially, few banks were contacted for the data; however, considering the sensitivity of the data requested, Banks could not provide the data. As a next option, relevant available datasets from Kaggle website were analysed and it was found that the identified datasets were not suitable for the research due to very limited attributes. Hence, a decision was made to create the synthetic data required for research.

Next sections describe the approach and steps followed to generate the data.

3.3.1 Synthetic Data Creation Approach

Figure 3.2 shows the synthetic data creation approach. The data entities includes customer, product and transactions. The key steps of the approach are as described below.

- **Identify attributes** - It starts with identifying the relevant data entities along with data attributes by studying the existing literature.

- **Define data population rules** - Next, define the data specifications such as data type, possible data values, importance of the data. Then, define the data creation rules which are used to create the data automatically. Define the suspicious transaction scenarios by studying the literature, especially the published documents by FIUs. The identified money laundering scenarios will be introduced in the data as a labelled training data.

- **Develop a utility** - Once the data attributes and data population rules are defined, a utility is developed to create the data automatically by using the rules. Post data creation by utility, further necessary data processing is done and the final dataset is produced that can be provided as an input to train and test the models.

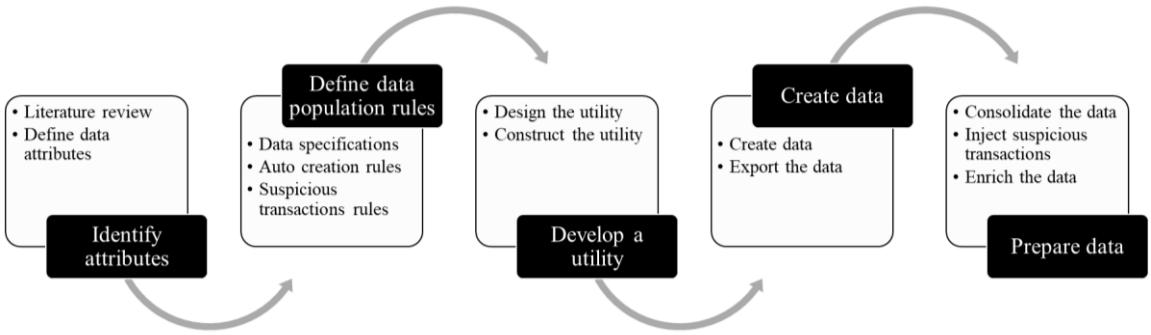


Figure 3.2 – Synthetic data creation methodology

- **Create data** – once the utility is developed, data would be generated by running the same. The generated data would be exported to an Excel file for further processing.
- **Prepare data** – the customer, transactions and accounts data in Excel format would be consolidated in a single file and enriched by adding suspicious money laundering transactions, labelling the suspicious and legitimate transactions and enriching the data. Finally, the data is saved as a CSV file that acts as an input to train and test the models.

3.3.2 Identify Data Attributes

Considering the context of detecting the suspicious money laundering transactions as Bank, the financial products and services provided by Banks are understood to determine the way customers interact with Bank, possibilities of money laundering transactions and relevant data attributes. After going through the relevant articles from literature as listed in **Table 2.3**, identified the type of data attributes used by researchers in their dataset. The list of attributes are identified as shown in **Figure 3.3** by studying the existing literature. The key data entities include customer, account and transactions. Some of the key attributes from the customer entity includes profession, income range, age, KYC status, risk rating which acts as key indicators while analysing the transactions. Account is a key entity which determines the source of transactions and helps to monitor the transaction behaviour for the account. Among the accounts such as savings, credit, loan, insurance; savings account is an important account where most of the cash transactions are initiated. Similarly, a transaction entity is at the centre that shows the flow of money, amount of

money, target account which helps to detect the transaction patterns and links. Hence, this thesis considered the customer, account and transaction as key data entities for data generation.

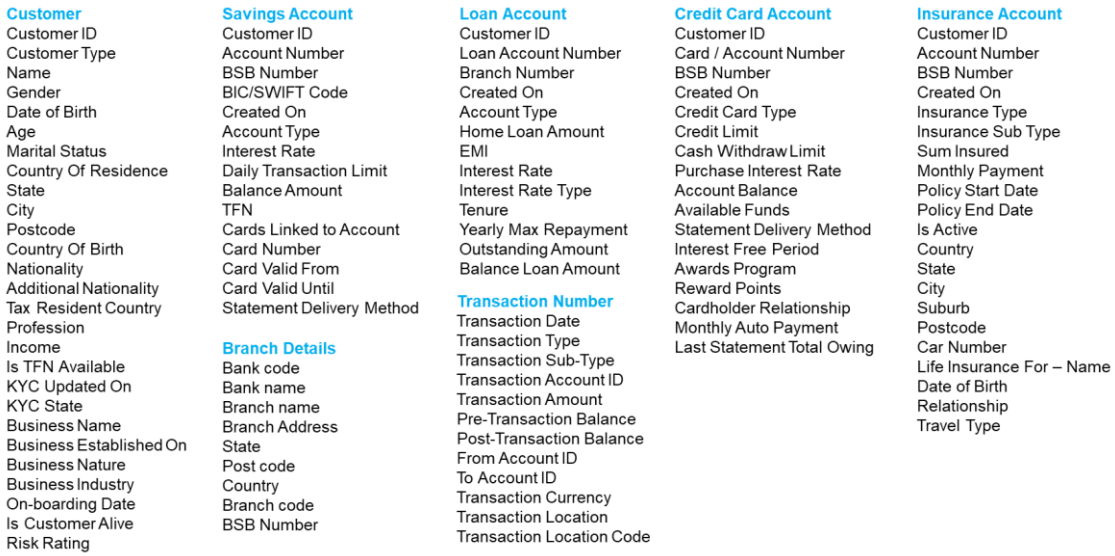


Figure 3.3 – Initial data attributes

For the purpose of research objective 1 and 2, considering the complexities around the transactions in the space of credit card, loan and insurance, the data attributes are kept limited to savings account, customer and transactions.

Table 3.1, **Table 3.2** and **Table 3.3** presents the final set of data attributes for customer, account and transactions respectively which are considered for research**.** The automatically generated data contains the strings; however, the same is converted into numeric values during data processing which can be understood by the machine learning models.

Table 3.1 – Customer demography data attributes

S. No.	Attribute	Data type	Constraint	Description
1	customerid	Number	Unique value	Customer number
2	customertypeid	Number	List value	Specifies if the customer is an individual or a business.
3	genderid	Number	List value	Gender. List values are – male, female, not specified
4	dateofbirth	Date	-	Date of birth
5	age	Number	-	Age
6	maritalstatusid	Number	List value	Marital state. List values are - single, married, divorced, widowed.
7	countryofresidenceid	Number	List value	Country of residence
8	stateid	Number	List value	State of residence
9	cityid	Number	List value	City of residence
10	postcode	Number	-	Post code of residence
11	birthcountryid	Number	List value	Birth country
12	nationalitycountryid	Number	List value	Nationality country
13	additionalnationalitycountryid	Number	List value	Additional nationality country
14	professionid	Number	List value	Profession. List values are - Managers, Professionals, Technicians and Trades Workers, Community and Personal Service Workers, Clerical and Administrative Workers, Sales Workers, Machinery Operators and Drivers, Labourers

S. No.	Attribute	Data type	Constraint	Description
15	incomeid	Number	List value	Income range. List values are – none, 1-18200, 18,201-45,000, 45,001-120,000, 120,000-180,000, 180,001 and over.
16	kycupdatedon	Date	-	KYC details update date
17	kycstateid	Number	List value	KYC state. List values are - expired, active.
18	riskratingid	Number	List value	Risk rating of customer (low, medium, high)
19	taxresidentcountryid	Number	List value	Tax resident country

Table 3.2 – Account data attributes

S. No.	Attribute	Data type	Constraint	Description
1	accountno	Number	Unique value	Account number
2	bsbno	Number	-	Bank state branch number
3	createdon	Date	-	Account opening date
4	accounttypeid	Number	List value	Account type. List values are – savings account, current account.
5	dailytranslimit	Number	-	Daily transaction amount limit
6	tfnid	Number	Unique value	Tax file number
7	statementdeliverymethodid	Number	List value	Specifies the statement delivery method. List values are – post, email
8	issuspicious	Number	-	0 indicates legitimate transaction and 1 indicates suspicious transaction

Table 3.3 – Transactions data attributes

S. No.	Attribute	Data type	Constraints	Description
1	transactiondt	Date	-	Transaction date
2	transactionno	Number	Unique value	Transaction reference number
3	transactionaccountid	Number	Unique value	
4	transactionamount	Number	-	Transaction amount
5	credit	Number	-	Specifics the amount credited in to account.
6	debit	Number	-	Specifies the amount debited from account.
7	transactiontypeid	Number	List value	Transaction type. List values are - Credit, Debit.
8	transactionsubtypeid	Number	List value	Transaction sub type. List values are – auto-debit, bill payment, cash deposit, cash withdrawal, international money transfer, online payment, overseas deposit, salary.
9	descriptionid	Number	List value	Predefined description string ID. List values are – broadband, car insurance, car servicing, cash deposit, cash withdrawal, CTP, domestic transfer, energy bill, gas bill, green slip, grocery, health insurance, income, Netflix, overseas deposit, overseas transfer, rent, shopping, telephone bill, transfer, transport card.
10	transactioncurrencyid	Number	List value	Transaction currency

S. No.	Attribute	Data type	Constraints	Description
11	transactionlocationtype	Number	List value	Specifics if the transaction is performed online or at branch. List values are – ATM, branch, online.
12	transactionlocationcode	Number	-	If the transaction is performed at branch, this represents the location code of a branch.
13	targetaccount	Number	-	Target account number
14	targetcountrycode	Number	-	Target account country code
15	bankcode	Number	-	Target bank code.
16	issuspicious	Number	-	Specifies the transaction label as legitimate or suspicious for training purpose.

3.3.3 Define Data Creation Rules

For the purpose of this thesis, the financial transaction dataset is decided to be created for a developed country. The customer data is generated as per the following rules:

- **Nationality** - Australian citizens and residents.
- **11 Cities** considered - Sydney, Melbourne, Brisbane, Perth, Adelaide, Gold Coast, Newcastle, Canberra, Sunshine Coast, Central Cost, and Wollongong.
- **Products** - savings account, home loan, vehicle loan, vehicle insurance.
- **Age group** - 25 to 60 years including male and females.
- **8 occupations** considered (as per NSW website) – Managers (1 male : 1 female), Professionals (1:1), Technicians and Trades Workers (2:1), Community and Personal Service Workers (1:1), Clerical and Administrative Workers (1:1), Sales Workers (1:1), Machinery Operators and Drivers (1:0), Labourers (1:0).
- **Income range** considered – 60,000 to 150,000 AUD annually.

Each customer is created with one savings account.

Transaction data is generated as per the following rules. Savings account will have the following transactions:

- **Credit transactions** – salary, cash deposits, incoming transfers
- **Debit transactions** – rent, home loan EMI, energy bills, phone bills, vehicle loan EMI, health insurance, vehicle rego, green slip, vehicle insurance, school fees, cash withdrawal, payment at counters, money transfer, shopping, travel and entertainment, subscriptions and bills.

In the case of developing countries, the transaction pattern would be different, where cash transactions would be more and of a higher amount.

3.3.4 Define Money Laundering Scenarios

The synthetic data generated so far is all legitimate transactions. In order to design the money laundering scenarios following sources are studied to understand the money laundering scenarios published by regulatory:

- Typologies and case studies (AUSTRAC 2022)
- Money laundering case studies (FinCEN 2022)

Table 3.4 shows the scenarios of suspicious money laundering transactions considered to create the labelled data and the same are injected into 298 accounts.

Table 3.4 – Suspicious money laundering transactions

S. No.	Scenario description	Number of records
1	Small deposits (< 5000) of money through ATM by multiple people into a single account (<10,000 per day) over a month. Then the same money is transferred in batches of 10,000 to 30,000 to multiple overseas account in different countries.	505
2	Small deposits (< 5000) of money through ATM by multiple people into a single account (<10,000 per day) over a month. Then the same money is used to buy luxurious item locally in the range of 10,000 to 90,000 (vehicles, gold, property etc).	168
3	Deposits of money from multiple overseas accounts from multiple countries and using the same money to buy luxurious items in the range of 10,000 to 90,000 AUD (vehicles, gold, property)	148
4	Deposits of money from multiple overseas accounts from multiple countries and withdraw the same through ATM over next couple of months in a small quantity in the range of 2000 to 4900 AUD.	108
5	Each month deposit a small sum of money in the range of 2,000 to 4,500 from ATM and next day transfer the deposited amount online to an account in a different local bank (but same account).	216

Table 3.5 shows the legitimate transactions scenarios that are similar to suspicious transactions, also called as overlapping transactions.

Table 3.5 – Legitimate overlapping transactions

S. No.	Scenario description	Number of records
1	Wire transfer of money to offshore accounts	100
2	Cash withdrawal from the account in the range of 2000 to 5000 AUD	100
3	Wire transfer of money from offshore account	50
4	Shopping in the range of 10,000 to 30,000 AUD	67

In the case of developing countries, the transaction pattern would be different and of large amounts. Since the legitimate transactions would also be highly cash transactions, it can have a big overlap on the actual money laundering transactions. This thesis has considered to create the transactions that are aligned with developed countries.

3.3.5 Design Synthetic Data Generation Software

The synthetic data generation software is written into .NET C# code where the modular functions are written to create a specific type of transactions. **Figure 3.4** presents the high level flow of auto-generation of data.

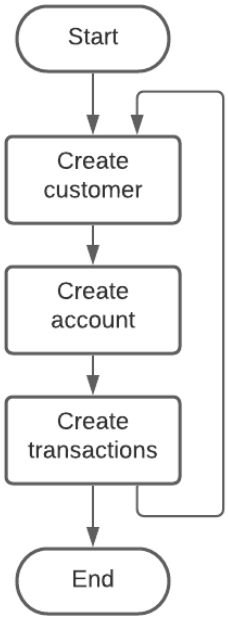


Figure 3.4 – Synthetic data generation flow

The utility runs the loop to create the customers based on the data rules defined for the customer. For each customer, one savings account is created. It is assumed all the customers are retail customers of a Bank having a savings account. Then the transactions on the savings account are generated for each month based on the transaction rules defined that includes credit and debit type of transactions. The transactions are generated for a period of approximately 12 months.

Transaction amount values are selected based on the combination of customer profile, type of transactions and a random amount value within the range. Example – electricity bill is generated based on the season and number of family members.

Fixed value transactions - There are some transaction values which are expected to be fixed for a certain period. Example – health insurance amount, rental payment, telephone bills, monthly subscriptions.

Annual transactions - The transactions also include some annual transactions such as – CTP, Green Slip, and Car Insurance.

Monthly income transactions – This includes monthly income transactions, fortnightly income transactions, and cash deposits.

Small amount transactions – this includes the transactions less than hundred dollars from a coffee buy to some grocery shopping.

High value transactions – this includes the transactions for shopping of articles or spending the money that varies from $500 to $5000 amount.

An attempt is made to ensure the transactions generated reflect the natural pattern of the transaction for any retail banking customer. **Figure 3.5** below presents the logic of generating the data for customers, accounts and transactions.

Pseudo code for synthetic data creation

```
for state in (Australia) {
   for city in (list of cities) {
      for profession in (list of professions) {
         for gender in (list of genders) {
            for martial_status in (list of marital status) {
               for age in (list of age groups) {
                  CreateCustomer();
                  CreateAccount();
                  CreateTransactions(){
                     CreateDebitTransactions(){
                        Income();
                        CashDeposits();
                        ReceiveMoney();
                     }
                     CreateCreditTransactions(){
                        Rent();
                        EnergyBills();
                        HealthInsurance();
                        TelephoneBills();
                        Boradband();
                        TVConnections();
                        Grocery();
                        Gas();
                        TransportBills();
                        Entertainment();
                        CarServicing();
                        CTP();
                        CarInsurance();
                        GreenSlip();
                        CashWithdraw();
                        TravelTickets();
                        MedicalBills();
                        Shopping();
                        RestaurantBills();
                     }
                  }
               }
            }
         }
      }
   }
}
```

Figure 3.5 – Synthetic data generation logic

The synthetic data generation logic is represented using scientific formulas as below:

Customer dataset is represented by 'c' as shown by equation **(3.1)**,

$$c = \sum_{j=1}^{M} f(s, ci, p, g, m, a)_j \quad \textbf{(3.1)}$$

where 'c' is a summation of the function of state (s), city (ci), profession (p), gender (g), marital status (m) and age (a). *M* represents total number of customers.

Account dataset is represented by 'a' as shown by equation **(3.2)**,

$$a = \sum_{j=1}^{M} c_j f(ty)_j \quad \textbf{(3.2)}$$

where 'a' is a list of accounts created for *M* customers. Account is a function of account-type (ty) and customer (c_j).

Transaction data for a single customer is shown by equation **(3.3)**,

$$t(a) = \sum_{j=1}^{M} f(d, cr, dt, \$, tp)_j a_j \quad \textbf{(3.3)}$$

where $t(a)$ is a list of transactions created for a single customer, which is a function of date (d), credit (cr), debit (dt), amount (\$), and type of transaction (ty).

A combined transaction dataset (T) containing customer, account and transaction information is represented by equation **(3.4)**,

$$T = \sum_{j=1}^{M} c_j a_j + t(a_j) \quad \textbf{(3.4)}$$

where T is a joined dataset of customer (c_j), accounts (a_j) and transactions ($t(a_j)$). The $t(a_j)$ represents the total transactions associated with account (a_j).

3.3.6 Generate Data Using Utility

Data is generated by running the utility. Following parameters are used to expand and ensure the balance in the dataset.

- Country – Add the country's name to create customer
- State – Add the state within the country
- City – Add the city with the state
- Profession – Add the profession
- Population ratio – Male-female ratio
- Gender – Male / female / do not want to say
- Marital status – Single / married / divorced / widowed
- Age – age bracket from 16 to 80
- Fixed transactions per month – the transactions with fixed amount and interval (example – health insurance, telephone bill, monthly instalments etc).
- Variable transactions per month – the transactions with variable amount and variable number of transactions for each customer. The amount is selected by capturing the random number from a fixed ranged of numbers for a specific transaction. The transaction date is generated using random function between a start and end date of a month. (Example – grocery transaction is generated between 1 to 5 times per month in a range of $10 to $250).
- Fixed transactions per year – this includes the transactions that are done once per year (Example – Vehicle insurance, registration, home insurance etc.).
- Variable transactions per month – this includes the transactions which are neither monthly nor annual transactions. (Example – holiday expenses, medical expenses, gift shopping, etc).

Appendix D shows the steps to generate the data, export it and inject suspicious transactions.

3.3.7 Define Data Enrichment Rules

Data enrichment is an essential step before inputting the data to the machine learning model. The enrichment includes several steps such as data cleansing, creating additional data from the existing data, organising the data, converting the data into appropriate formats. Machine learning classifiers do not accept the synthetically generated data as-is, because it contains the dates and text data. Hence, the following data enrichment is performed.

- **Covert text data into numeric**

Machine learning classifiers don't accept the strings hence all strings are assigned the stringcode which is a numeric value. The text field is then replaced by associated stringcode. The equation **(3.5)** presents the string conversion formula that converts text into numeric value. The text value is converted only once. If the text value repeats in the records then the earlier generated numeric value will be used.

$$s = \sum_{j=1,i=0}^{M,N} ConvertStringToNumeric(i)_j \qquad \textbf{(3.5)}$$

's' represents the dataset with string values converted into numeric values, 'j' represents the record number, and *'i'* represents the column number with the record.

- **Convert date into numeric**

Date values are also not accepted by machine learning classifier in date format, hence they are converted into numeric value using Excel facility by converting date cell into numeric value. The equation **(3.6)** presents the date conversion formula that converts a date into numeric value. The date value is converted only once. If the date value repeats in the records then the earlier generated numeric value will be used.

$$d = \sum_{j=1,i=0}^{M,N} ConvertDateToNumeric(i)_j \qquad \textbf{(3.6)}$$

'd' represents the dataset with date values converted into numeric values, 'j' represents the record number, and *'i'* represents the column number with the record.

- **Add noise in the dataset**

Synthetically generated data comes up with a specific pattern with all values are filled logically well. However, practically this may not be a scenario, the real data may have a large variation in terms of transaction value and frequency of transactions.

Supervised models learn from the training data provided and having a known pattern can lead to overfitting the model. Hence, to avoid the overfitting the model, the datasets is injected with a noise to break that pattern, make the data more generic and make it a bit more difficult for the model to learn and predict. The issue with overfitting is, the model works well with the dataset on which it is trained, but if a different dataset is provided then it fails to predict. The noise is added to data by creating ~5000 similar records from existing records, and ~5000 records are updated with a null or 0 value randomly.

3.4 Objective 1 – Detect Suspicious Transactions

Objective 1 of this study is to design an effective deep learning model to accurately detect the suspicious money laundering transactions. CNN is considered to address this objective.

The literature (Chen et al. 2018b) shows frequent use of Decision Tree, SVM, Naïve Bayes, RF, Logistic Regression, XGBoost, Clustering machine learning classifiers to detect suspicious money laundering transactions. Three classifiers are chosen from the list based on the frequency of usage and the accuracy of the predictions for comparing the results of CNN classifiers. These classifiers are – RF, SVM and XGBoost.

3.4.1 Convolutional Neural Network

Deep Learning (Lecun et al. 2015) is a subset of Machine Learning. Deep learning is inspired by, the way brain functions, the way neurons get activated and communicate to other neurons leading to an action or thought (Miyake and Fukushima 1984). CNN is one of the method of deep learning which is heavily and successfully used in computer vision domain (Sabokrou et al. 2021) for image analysis (Palraj and Kalaivani 2021), classification, and identification; and also for speech recognition (Li et al. 2021), natural

language processing (Roshanzamir et al. 2021) and time series data (Kim et al. 2021). CNN is also one of the most researched algorithm to address the challenges around the computing needs, which has helped to refine and make the algorithm more efficient. Automatic feature extraction capability of CNN is also another attraction for researchers with complex data (Miyake and Fukushima 1984).

In this work, CNN classifier is identified to detect suspicious transactions from a tabular transaction data. Typically, CNN design consists of an input layer, multiple hidden layers and an output layer (Fan et al. 2019). **Figure 3.6** shows the architecture of proposed CNN method that includes Conv1D layer, Normalisation, Flatten, Dropout and Dense layers.

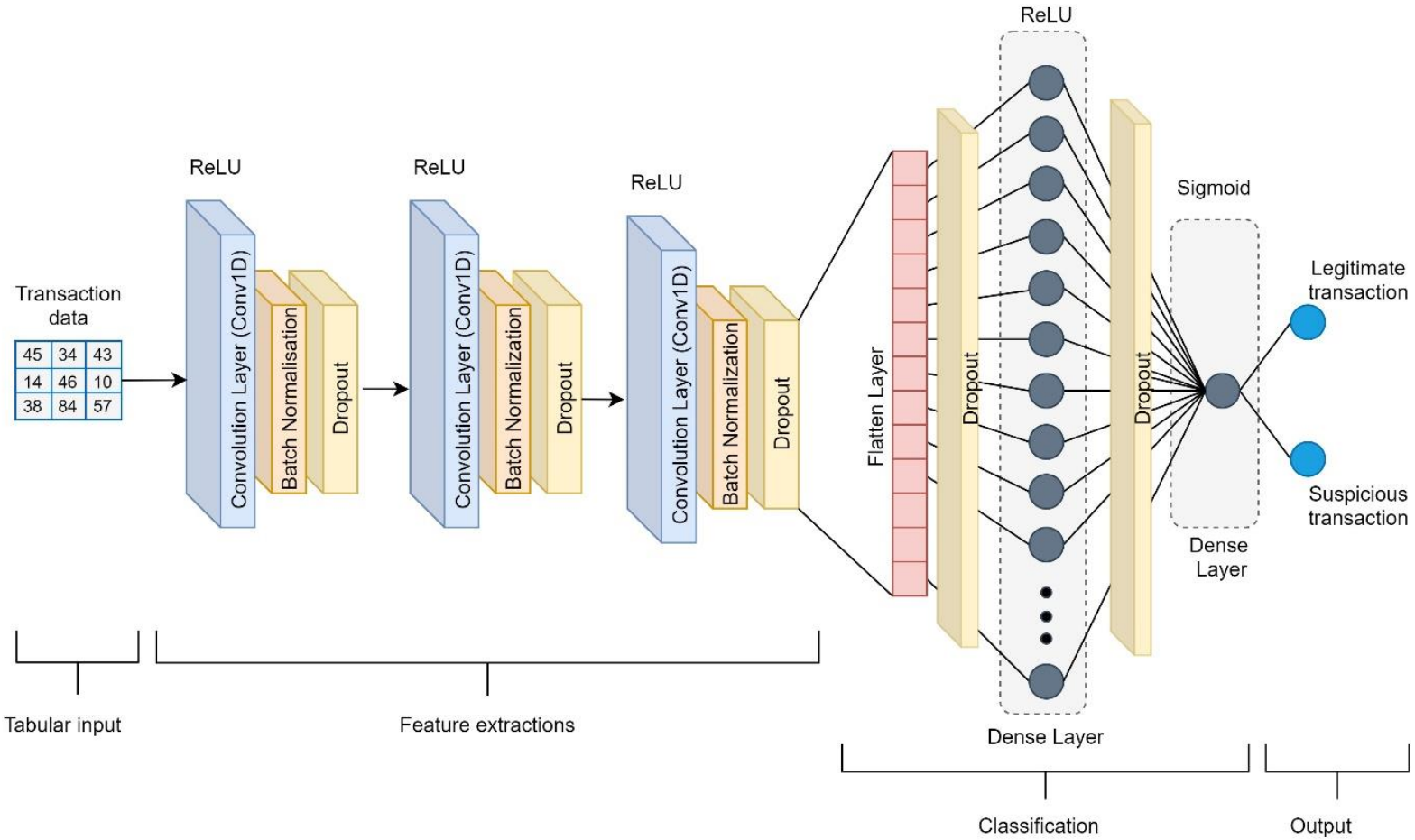


Figure 3.6 – CNN architecture

The proposed CNN architecture has used the 5 different types of layers from a Sequential class published in Keras library (Chollet 2015) that is built on top of TensorFlow (Team 2015). These layers are described as below:

Convolution layer – this layer performs the convolution operation on the input data provided, produces the output in the vector format and passes it to the next layer (Miyake and Fukushima 1984), (Schmidhuber 2015). There are various types of convolution layers such as Conv1D, Conv2D, Conv3D (Chollet 2015). Conv1D layer usually takes the input with one dimension such as time-series data, voice or text. The Conv1D kernel convolutes in one direction. Conv2D layer takes the input with 2 dimensions such as image; and the Conv2D kernel moves in 2 directions. Conv3D kernel takes the inputs such as video, where kernel convolutes in 3 directions. Other than the image for each frame, there is another axis as time with corresponding frame. For the purpose of this study, we have used Conv1D type of layer with a tabular dataset. The shape of data input and output from this layer is 1 dimensional. ReLU activation function (Nair and Hinton 2010) is used with Conv1D layer.

Batch normalisation layer – this layer applies the normalisation for input data that helps to maintain the output mean close to 0 and standard deviation close to 1 (Ioffe and Szegedy 2015). This layer works differently during training and predicting. During training, the layer normalises the output using the mean and standard deviation of the current batch of input. During prediction, the layer normalises the output using the moving average of mean and standard deviation of the batches it has seen during training (Chollet 2015).

Dropout layer – dropout layer randomly sets the input to 0 during training at a frequency of rate defined, that helps to prevent overfitting issue and improves the model generalisation power (Srivastava et al. 2014; Chollet 2015). Dropout layer is applicable only during training (Chollet 2015).

Flatten layer – flatten layer flattens the input, meaning coverts the multi-dimensional inputs into a single dimension array which are passed to Dense layer.

Dense layer – dense layer receives the input from each neurons of the previous layer and uses to change the dimension of vector (Chollet 2015; Huang et al. 2017). We are using

ReLU (Nair and Hinton 2010) as an activation function for the hidden dense layer and sigmoid (Han and Moraga 1995) activation function for the output dense layer.

3.4.2 Random Forest

RF is a decision tree based supervised machine learning classifier frequently used for classification and regression analysis. The classification prediction is determined by a class selected by majority trees. RF uses an ensemble learning technique that combines (by taking mean or average of) the outcome of many decisions trees for prediction by training using the Bagging method. The number of trees is one the key factor for precise prediction. One of the drawback of decision tree classifier is overfitting of dataset, which RF classifier addresses by increasing the precision of predictions. It is also one of the simplest algorithm that doesn't need much of hyper-parameter tunning.

Literature provides several references where RF classifier is used in AML domain to cite few - (Magomedov et al. 2018), (Badal-Valero et al. 2018b). RF classifier is used to identify suspicious transactions and detect the money laundering groups by setting up the number of trees as hundred and remaining parameters as default (Savage et al. 2017b).

3.4.3 Support Vector Machine

SVM is a classifier from the supervised learning category that predicts the results based on the training data. The predictions are either classification, regression or clustering. SVM constructs one or more hyperplanes that are used for classification, regression or outlier detection. The algorithm finds the points that are closer to the hyperplane on both the classes, these points are called as support vectors. Then the distance between support vectors and line is calculated which is called as margin. The goal of SVM is to maximize the margin to find the optimal hyperplane that can predict the classes as accurately as possible.

SVM classifier (kernel and polynomial kernels) is used to detect the money laundering groups by analysing the transactions (Savage et al. 2017b), a research that was funded by AUSTRAC. There are few other researchers who have used SVM in money laundering domain - (Feng et al. 2019), (Zhang and Trubey 2019), (Guevara et al. 2020b), (Shokry et al. 2020).

3.4.4 Extreme Gradient Boosting

XGBoost is an optimized open source library (Chen and Guestrin 2016). XGBoost classifier is an ensemble method based on the decision tree. The core feature of XGBoost is its ability of parallelism within a single tree that boosts the performance and speed. This classifier can utilize the full potential of today's multi-core CPUs. The decision trees of XGBoost classifiers ends up with real-value in each leaf, regardless of whether the classification or regression. Real-value scores are then used for classification if it is a classification problem. XGBoost use is demonstrated to predict the suspicious transactions by calculating the probably of the transaction being suspicious (Jullum et al. 2020). Another researcher has used XGBoost classifier to detect illicit accounts involved in money laundering over ethereum blockchain (Farrugia et al. 2020).

3.4.5 Evaluation and Performance Metrics

In order to measure the performance of classifiers used to detect suspicious transactions, we considered the following metrics:

True Positives (TP) and **True Negatives (TN)** are the transactions that are correctly classified (Sammut and Webb 2010h, 2010g). More close the number of predicted true positives and true negatives with actual true positives and true negatives, the better model it is.

False Positives (FP) and **False Negatives (FN)** are the transactions that are incorrectly classified (Sammut and Webb 2010e, 2010d). More close the number of predicted false positives and false negatives with 0, the better model it is. With respect to money laundering domain, false positive and false negative has a specific importance. False positive indicates that the model has predicted the legitimate transaction as a money laundering transaction and this results into AML officer manually reviewing it and eventually closing it – this increases the operational cost for organisation. On the other hand, false negative indicates that the model has predicted money laundering transaction as legitimate and this results into missing the actual money laundering transaction leading to allowing money launderers to use the financial system, which is far more dangerous and costly than the operational cost of efforts spent on reviewing the false positives. More false negatives is an indication of system being not effective that can lead the organisation

into non-compliance of effective AML controls, risks of a heavy penalty by regulatory, damage to brand and corporate citizenship.

Accuracy is the percentage of transactions correctly classified (Sammut and Webb 2010a) as defined in equation **(3.7))**. Accuracy is one of the best metric to understand and measure the model performance, but is not suitable when the data is highly imbalanced. In finance data and fraud scenarios, the true negatives are way higher than the false positives and false negatives together; and it undermines the impact of false positives and false negatives. Hence, we considered accuracy as a measure to see the model performance as overall, but this is not our key measurement criteria.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad \textbf{(3.7)}$$

Precision is the percentage of predicted positives that are correctly classified (Ting 2010a, 2010b) as defined in equation **(3.8)**. Usually if the cost of false positives is higher than false negative, then precision is better a metric. In money laundering scenario, this is important metrics as it deals with the operational cost of effort required for reviewing false positives.

$$precision = \frac{TP}{TP + FP} \quad \textbf{(3.8)}$$

Recall is the percentage of actual positives that are correctly classified (Sammut and Webb 2010f; Ting 2010b) as defined in equation **(3.9)**. Usually, if the cost of false negatives is higher than false positives, then recall is a better metric. In money laundering scenario, this is a highly important metric as it involves high risk of non-compliance, penalties and brand image due to false negatives or missing out on the true money laundering transactions.

$$recall = \frac{TP}{TP + FN} \quad \textbf{(3.9)}$$

F1 score combines the precision and recall metric, and determines the harmonic mean to measure the performance of the classifier (Sammut and Webb 2010c) as defined in equation **(3.10)**. Type-I error is represented by the number of false positives returned by a model and type-II error is represented by the number of false negatives returned by a model. Error importance depends on the domain of the classification problem. In case of AML we gave higher importance to type-II error after consulting with AML experts in the industry, type-I error means more alerts and more operational efforts to investigate the alerts, but type-II error indicates the system is not effective in identifying true suspicious transactions which is a bigger risk from compliance point of view as well as allowing the launderers to exploit the financial system for their benefits. Hence, the model is required to produce as less false negatives as possible to improve suspicious transaction detection. Type-II error with high importance implies that recall metric is more important than precision, hence we have considered $f_\beta$score metric calculation with $\beta = 3$ to reflect higher importance to recall metric. Generally the default value of β is 1, which represents both the classes in the classification problem has same weightage. We have used the following formula to calculate the f-score with $\beta = 3$, considering 3 times weightage for false negative results over false positives.

$$f_\beta score = (1 + \beta^2) \times \frac{precision \times recall}{(\beta^2 \times precision) + recall} \quad \textbf{(3.10)}$$

AUC score is used to identify the models capability to distinguish between legitimate and suspicious class of transactions (Sammut and Webb 2010b; Flach 2010). AUC score represents the area under the curve plotted using true positive rate and false positive rate, as defined in equation **(3.11)**. The more area under the curve, better the model it is.

$$auc\ score = \frac{1}{2}\left(1 + \frac{TP}{TP + FN} - \frac{FP}{FP + TN}\right) \quad \textbf{(3.11)}$$

3.5 Objective 2 – Explain Predictions

Objective 2 of this thesis is to design a method to explain the reasoning behind the predictions made by the chosen deep learning method which is CNN.

At the current stage of technological advancements and AI/ML adoption in the banking industry, when it comes to selecting the classifiers for prediction, it is always a trade-off between accuracy and interpretability. Interpretability means how accurately a model can associate a cause to an effect. Better interpretability leads to better explainability of a model. Here, explainability means how well the interpretability can justify the predictions. Domain such as weather prediction may value accuracy more than the model interpretability, but the domains where the post-decision stakes are high, the model interpretability is valued more to enable the adoption. The domain such as medical (Tjoa and Guan 2021), regulatory compliance (Huynh et al. 2021; Hall et al. 2019) and criminal investigations (Hepenstal et al. 2021) require the model to be interpretable or explainable. AML is one of such domain where model interpretability is critical (Kuiper et al. 2021).

The transparency, fairness and accountability are paramount in finance industry. XAI can help in bringing the fairness in decision making while investigating the alerts (Fritz-Morgenthal et al. 2022). To understand the perspective of application of XAI in finance sector from supervisory authority and reporting entities, (Kuiper et al. 2022) conducted a study using semi-structured interviews with 3 banks and 2 supervisory authorities to examine 3 use cases - AML, credit risk and consumer credit. The study found disparity in a view of bank and supervisory authorities on whether XAI should be used or not in the given scenarios. Banks are leveraging AI capability wherever possible, whereas SAs do not prescribe of using any particular technology, and they are fine as long as they receive the reports with sufficient details. In AML use case, both the banks are using AI to reduce the noise (false positives) and improve the detection. One Bank indicated that they have started using SHAP XAI technique on deduplication and classification algorithm to explain the fraudulent activities detected in corresponding banking transactions, however, no reference of such combined method found in literature.

In Banks, the suspicious transactions alerts raised by AML system always go through an investigation by AML officer, hence the interpretability can help the officer investigate

the suspicious transactions more effectively. To gain the benefits of deep learning models which are black-box in nature, it is important to know how decisions are made by such models. Recently several post-hoc methods have been developed to explain the opaque or black box models such as SHAP (Lundberg and Lee 2017), LIME (Ribeiro et al. 2016) , which applies yet another model on top of black-box model and tweak the input to see the impact on output values. That helps to gain the understanding of feature importance considered by the black-box models for making the decisions. Some researchers have a different view that says when the decision stakes are high, interpretable methods should be preferred over post-hoc approach for interpretability and future research should be focused on creating interpretable models (Rudin 2019).

For the purpose of this research, a model-agnostic XAI method SHAP (Lundberg and Lee 2017) was chosen to explain the predictions made by CNN classifier. SHAP is based on the coalitional game theory (Shapley 1953), where feature values are considered as team players; dataset is considered as a team and the game result is considered as a prediction. Each feature contributes for predicting the result. SHAP quantifies the contribution made by each feature value in the prediction made by the model. Here, each feature value represents one value within a single record. The core idea of SHAP is to reverse-engineer and calculate the impact of each feature value on the target predicated value.

Shapley value explanation is represented as a linear model and an additive feature attribution method (Molnar 2019). SHAP (Lundberg and Lee 2017) uses the earlier proposed explanation methods DeepLIFT (Shrikumar et al. 2017) and LIME (Ribeiro et al. 2016) to explain the prediction for original model $f(z)$ as shown in equation **(3.12)**:

$$f(z) = g(z') = \emptyset_0 + \sum_{j=1}^{M} \emptyset_j Z'_j \qquad \textbf{(3.12)}$$

Where, $f(z)$ represents the CNN model in this case, z is the input parameters or transaction data, $g(z')$ represents the explanation model, $z' \in \{0,1\}^M$ is the coalition vector, M is the maximum coalition size and $\emptyset_j \in$ R is a feature j attribution.

SHAP provides multiple explainers such as Tree Explainer, Gradient Explainer, Deep Explainer, Kernel Explainer (Lundberg 2018). A DeepExplainer which is an enhanced

version of DeepLIFT (Shrikumar et al. 2017) algorithm, from SHAP library (Lundberg 2018) was used to generate the Shapley values for CNN classifier, and then bring out the interpretations of predictions using various plots supported by the library.

3.6 Software for Model Implementation

This section describes the software and tools used for modelling and synthetic data creation.

3.6.1 Modelling Software

Table 3.6 provides a list of software tools and libraries used for developing the code for training, evaluating and testing the CNN, RF, SVM, XGBoost and SHAP models.

Table 3.6 – Modelling software used

Software	Description	Use
Python 3.0 (Foundation 1991)	Python is a general purpose programming language.	The code for machine learning models is written in python language.
Jupyter Notebook 6.0.3	Jupyter Notebook is an integrated development environment for writing python code.	Jupyter Notebook is used for writing the modelling code, training and testing the machine learning model.
Keras library (Chollet 2015)	Keras is built on top of TensorFlow 2.0. Keras is a library APIs for deep learning framework.	Keras library is used to develop, train and test the CNN model.
SHAP library (Lundberg 2018)	SHAP is a library based on game theory used to explain the decisions made by machine learning models.	SHAP is applied on CNN model to explain the predictions made by CNN.

3.6.2 Data Creation Software

Table 3.7 presents the software and tools are used for generating the synthetic data for research.

Table 3.7 – Data creation software used

Software	Description	Use
Microsoft C#	C# is a object oriented programming language developed by Microsoft.	Data generation code is written to implement the rules and create the data.
PostgreSQL	PostgreSQL is an open source relational database management system.	Database tables were created to store the generated data.
Microsoft Excel	Excel is a spreadsheet developed by Microsoft.	After exporting the data from PostgreSQL into Excel, initial data pre-processing was done before converting the data into CSV.

3.7 Summary

Finance transaction data entities, attributes and data creation rules are defined assuming the data for the retail banking customers. A Microsoft .NET based software is developed to auto-create the synthetic finance data that includes customers, accounts and transactions. Basic parameters are defined which are taken as an input to generate the synthetic data. The generated synthetic data is enriched to ensure the dataset is numeric and also ingested with some noise. Further suspicious transaction scenarios are defined and relevant transactions are ingested in the training dataset. Then the transactions are marked as legitimate and suspicious transactions which is used as a labelled data by machine learning models.

A methodology to address objective 1 of thesis is designed by defining a 1D CNN deep learning technique to detect suspicious money laundering transactions, and defining associated hyper parameters to experiment with. To compare the prediction performance of CNN, following traditional machine learning classifiers defined – RF, SVM and

XGBoost; and following performance metrics are chosen – confusion matrix, precision, recall and $f_\beta score\ with\ \beta = 3$.

A methodology to address Objective 2 of the thesis is designed by defining a method to apply SHAP XAI technique on CNN classifier which generates the Shapley values for each value in the dataset that helps to determine the positive or negative influence on the prediction. Finally, the chapter provides the software's and tools used for synthetic data creation and modelling.

4 RESULTS AND DISCUSSION

4.1 Introduction

The synthetic data was produced using a custom developed synthetic data generator software by authors and was refined further during experiments. The synthetic dataset contains customers, accounts, legitimate transactions and suspicious money laundering transactions. The code was written to predict the suspicious transactions using a CNN classifier. The CNN model was trained and tested on the produced dataset, and validated the performance against results produced by RF, XGBoost and SVM classifiers for the same dataset. To explain the predictions made by the CNN model, the code was written to use the DeepExplainer class of SHAP library to produce the Shapley values and explain the predictions made by CNN classifier at individual record level. This section provides the results of all experiments.

4.2 Synthetic Data

The input data for detecting suspicious transactions in Banks includes customer details, account details and transactions. The transactions are classified in two classes – legitimate and suspicious. By nature, the transaction data is a highly imbalanced data, where majority class is legitimate transactions and minority class is suspicious transactions. Considering the same principle, the highly imbalanced data was produced with a ratio of 95.64 to 4.36 percentage.

Table 4.1 provides the specifications of the data produced and used for training and testing the models. **Table 4.2** presents a sample of customer data produced by the synthetic data generator software. The data value types are string, numeric, and date. Finally, the dataset is covered into numeric dataset. Please refer to section **3.3** for more details on data specifications. **Table 4.3** presents the accounts data produced by the software. The dataset contains the strings, numbers and date type of values which are converted into a numeric dataset.

Table 4.4 presents the sample of transactions generated for each account.

Table 4.1 – Dataset specifications

Parameter	**Value**
Number of customers	442
Number of accounts	442
~Number of transactions per customer	184
~Time period of transactions	12 months
Total number of transactions	92824
Labelled suspicious transactions	4054
Labelled legitimate transactions	88770
Number of labels	2 labels with value 0 and 1. 0 – Legitimate transaction 1 – Suspicious transaction
Total data attributes before merging customer (20), account (15) and transaction (13) table	48 attributes
Total data attributes after merging customer, account and transaction table (Note: Columns representing foreign key are not required in merged dataset, hence they are removed.)	43 attributes (this includes isSuspicious column indicating if the row represents a suspicious or legitimate transaction, and isNoise column indicating if the row is injected with noise) **Note:** isNoise column is used only for administration purpose to know the noisy records.

Table 4.2 – Raw customer demography data sample

Customer ID	Customer Type	Customer Name	Gender	Date Of Birth	Age	Marital Status	CountryOfResidence	State	City	Post Code	Birth Country	Profession	Income	Is TFN Available	KYC Updated On	KYC State	On Boarding Date	Is Customer Alive	Risk Rating	Tax Resident Country
20000001	Individual	Theodore Nguyen	Male	2/10/1995	25	Single	Australia	New South	Sydney	2000	Australia	Managers	None	Yes	1/06/2019	Active	1/11/2017	Yes	Low	Australia
20000002	Individual	Lucas Brown	Male	15/06/1990	30	Single	Australia	New South	Sydney	2000	Australia	Managers	None	Yes	4/01/2019	Active	19/11/2017	Yes	Low	Australia
20000003	Individual	Jacob Williams	Male	17/01/1985	35	Single	Australia	New South	Sydney	2000	Australia	Managers	None	Yes	8/11/2019	Active	9/11/2017	Yes	Low	Australia
20000004	Individual	Samuel Ryan	Male	20/08/1980	40	Single	Australia	New South	Sydney	2000	Australia	Managers	None	Yes	12/03/2019	Active	24/11/2017	Yes	Low	Australia
20000005	Individual	Benjamin Martin	Male	9/02/1975	45	Single	Australia	New South	Sydney	2000	Australia	Managers	None	Yes	20/01/2019	Active	14/05/2017	Yes	Low	Australia
20000006	Individual	Benjamin Walker	Male	6/09/1970	50	Single	Australia	New South	Sydney	2000	Australia	Managers	None	Yes	16/08/2019	Active	1/03/2017	Yes	Low	Australia
20000007	Individual	Carter Nguyen	Male	26/06/1965	55	Single	Australia	New South	Sydney	2000	Australia	Managers	1 - 18200	Yes	14/03/2019	Active	27/03/2017	Yes	Low	Australia
20000008	Individual	Elijah Wilson	Male	10/09/1960	60	Single	Australia	New South	Sydney	2000	Australia	Managers	1 - 18200	Yes	8/10/2019	Active	1/04/2017	Yes	Low	Australia
20000009	Individual	Jacob Thompson	Male	23/02/1990	30	Married	Australia	New South	Sydney	2000	Australia	Managers	None	Yes	14/04/2019	Active	8/07/2017	Yes	Low	Australia
20000010	Individual	Benjamin Nguyen	Male	1/05/1985	35	Married	Australia	New South	Sydney	2000	Australia	Managers	None	Yes	20/01/2019	Active	17/11/2017	Yes	Low	Australia
20000011	Individual	Oliver Martin	Male	11/02/1980	40	Married	Australia	New South	Sydney	2000	Australia	Managers	None	Yes	6/08/2019	Active	16/02/2017	Yes	Low	Australia
20000012	Individual	Arthur Robinson	Male	13/09/1975	45	Married	Australia	New South	Sydney	2000	Australia	Managers	None	Yes	4/04/2019	Active	8/02/2017	Yes	Low	Australia
20000013	Individual	Benjamin Walker	Male	15/01/1970	50	Married	Australia	New South	Sydney	2000	Australia	Managers	None	Yes	3/06/2019	Active	5/05/2017	Yes	Low	Australia
20000014	Individual	Arthur Smith	Male	6/05/1965	55	Married	Australia	New South	Sydney	2000	Australia	Managers	1 - 18200	Yes	18/08/2019	Active	12/09/2017	Yes	Low	Australia
20000015	Individual	Elijah Ryan	Male	27/06/1960	60	Married	Australia	New South	Sydney	2000	Australia	Managers	1 - 18200	Yes	19/09/2019	Active	5/11/2017	Yes	Low	Australia
20000016	Individual	Lincoln White	Male	14/03/1990	30	Seperated	Australia	New South	Sydney	2000	Australia	Managers	None	Yes	23/10/2019	Active	4/09/2017	Yes	Low	Australia
20000017	Individual	Alexander Kelly	Male	19/09/1985	35	Seperated	Australia	New South	Sydney	2000	Australia	Managers	None	Yes	16/02/2019	Active	6/06/2017	Yes	Low	Australia
20000018	Individual	Samuel Johnson	Male	7/01/1980	40	Seperated	Australia	New South	Sydney	2000	Australia	Managers	None	Yes	11/04/2019	Active	10/04/2017	Yes	Low	Australia
20000019	Individual	Elijah Johnson	Male	6/08/1975	45	Seperated	Australia	New South	Sydney	2000	Australia	Managers	None	Yes	27/11/2019	Active	21/05/2017	Yes	Low	Australia
20000020	Individual	Cooper Thompson	Male	25/02/1970	50	Seperated	Australia	New South	Sydney	2000	Australia	Managers	None	Yes	3/10/2019	Active	14/08/2017	Yes	Low	Australia
20000021	Individual	Ethan Kelly	Male	8/11/1965	55	Seperated	Australia	New South	Sydney	2000	Australia	Managers	1 - 18200	Yes	24/08/2019	Active	20/06/2017	Yes	Low	Australia
20000022	Individual	Theodore Kelly	Male	9/10/1960	60	Seperated	Australia	New South	Sydney	2000	Australia	Managers	1 - 18200	Yes	19/05/2019	Active	3/11/2017	Yes	Low	Australia
20000023	Individual	Benjamin Anderson	Male	15/01/1990	30	Divorced	Australia	New South	Sydney	2000	Australia	Managers	None	Yes	14/06/2019	Active	11/02/2017	Yes	Low	Australia
20000024	Individual	Michael Taylor	Male	11/09/1985	35	Divorced	Australia	New South	Sydney	2000	Australia	Managers	None	Yes	26/07/2019	Active	11/10/2017	Yes	Low	Australia
20000025	Individual	Eli Walker	Male	17/09/1980	40	Divorced	Australia	New South	Sydney	2000	Australia	Managers	None	Yes	20/11/2019	Active	1/10/2017	Yes	Low	Australia

Table 4.3 – Raw accounts data sample

CustomerID	AccountNo	BSBNo	BISCode	CreatedOn	Account TypeID	Account Type	InterestRate	Daily Transaction Limit	Balance Amount	TFN	Cards Linked To Account	Card Valid From Date	Card Valid Until Date	Statement Delivery Method
20000002	10200001	201106	BANKAU5D	11/19/2017	2	Saving	1	5000	0	Yes	1	11/19/2017	11/19/2020	Online
20000003	10200002	201106	BANKAU5D	11/09/2017	2	Saving	1	5000	0	Yes	1	11/09/2017	11/09/2020	Online
20000004	10200003	201106	BANKAU5D	11/24/2017	2	Saving	1	5000	0	Yes	1	11/24/2017	11/24/2020	Online
20000005	10200004	201102	BANKAU5D	05/14/2017	2	Saving	1	5000	0	Yes	1	05/14/2017	05/14/2020	Online
20000006	10200005	201102	BANKAU5D	3/01/2017	2	Saving	1	5000	0	Yes	1	3/01/2017	3/01/2020	Online
20000007	10200006	201102	BANKAU5D	03/27/2017	2	Saving	1	5000	0	Yes	1	03/27/2017	03/27/2020	Online
20000008	10200007	201107	BANKAU5D	4/01/2017	2	Saving	1	5000	0	Yes	1	4/01/2017	4/01/2020	Online
20000009	10200008	201104	BANKAU5D	7/08/2017	2	Saving	1	5000	0	Yes	1	7/08/2017	7/08/2020	Online
20000010	10200009	201107	BANKAU5D	11/17/2017	2	Saving	1	5000	0	Yes	1	11/17/2017	11/17/2020	Online
20000011	10200010	201104	BANKAU5D	02/16/2017	2	Saving	1	5000	0	Yes	1	02/16/2017	02/16/2020	Online
20000012	10200011	201101	BANKAU5D	2/08/2017	2	Saving	1	5000	0	Yes	1	2/08/2017	2/08/2020	Online
20000013	10200012	201107	BANKAU5D	5/05/2017	2	Saving	1	5000	0	Yes	1	5/05/2017	5/05/2020	Online
20000014	10200013	201108	BANKAU5D	9/12/2017	2	Saving	1	5000	0	Yes	1	9/12/2017	9/12/2020	Online
20000015	10200014	201105	BANKAU5D	11/05/2017	2	Saving	1	5000	0	Yes	1	11/05/2017	11/05/2020	Online
20000016	10200015	201107	BANKAU5D	9/04/2017	2	Saving	1	5000	0	Yes	1	9/04/2017	9/04/2020	Online
20000017	10200016	201108	BANKAU5D	6/06/2017	2	Saving	1	5000	0	Yes	1	6/06/2017	6/06/2020	Online
20000018	10200017	201103	BANKAU5D	4/10/2017	2	Saving	1	5000	0	Yes	1	4/10/2017	4/10/2020	Online
20000019	10200018	201101	BANKAU5D	05/21/2017	2	Saving	1	5000	0	Yes	1	05/21/2017	05/21/2020	Online
20000020	10200019	201108	BANKAU5D	08/14/2017	2	Saving	1	5000	0	Yes	1	08/14/2017	08/14/2020	Online
20000021	10200020	201107	BANKAU5D	06/20/2017	2	Saving	1	5000	0	Yes	1	06/20/2017	06/20/2020	Online
20000022	10200021	201103	BANKAU5D	11/03/2017	2	Saving	1	5000	0	Yes	1	11/03/2017	11/03/2020	Online
20000023	10200022	201102	BANKAU5D	2/11/2017	2	Saving	1	5000	0	Yes	1	2/11/2017	2/11/2020	Online
20000024	10200023	201103	BANKAU5D	10/11/2017	2	Saving	1	5000	0	Yes	1	10/11/2017	10/11/2020	Online
20000025	10200024	201107	BANKAU5D	10/01/2017	2	Saving	1	5000	0	Yes	1	10/01/2017	10/01/2020	Online
20000026	10200025	201103	BANKAU5D	08/23/2017	2	Saving	1	5000	0	Yes	1	08/23/2017	08/23/2020	Online

Table 4.4 – Raw transactions data sample

Transaction No	Transaction Date	Transaction TypeID	Transaction SubTypeID	Transaction AccountID	Transaction Amount	Credit	Debit	Transaction CurrencyID	Transaction LocationID	Transaction LocationCode	Description ID	Txn Suspicious
2771	42923	25	128	10204497	5416	5416	0	136	140	222	1	0
2772	42923	26	131	10204497	50	0	50	136	140	222	5	0
2773	42923	26	132	10204497	16	0	16	136	140	222	13	0
2774	42923	26	134	10204498	1168	0	1168	136	140	222	16	0
2775	42923	26	131	10204481	80	0	80	136	140	222	12	0
2776	42923	26	133	10204869	200	0	200	136	140	222	18	0
2777	42923	26	134	10204917	487	0	487	136	140	222	17	0
2778	42923	26	131	10204908	80	0	80	136	140	222	12	0
2779	42923	25	122	10201968	7800	7800	0	136	142	168	19	1
2780	42923	26	131	10204921	437	0	437	136	140	222	4	0
2781	42923	26	131	10204900	50	0	50	136	140	222	6	0
2782	42923	26	131	10204895	80	0	80	136	140	222	12	0
2783	42923	26	132	10204851	1784	0	1784	136	140	222	2	0
2784	42923	26	134	10204851	510	0	510	136	140	222	15	0
2785	42923	26	132	10204857	90	0	90	136	140	222	7	0
2786	42924	26	132	10204857	315	0	315	136	140	222	3	0
2787	42924	26	131	10204851	20	0	20	136	140	222	6	0
2788	42924	26	132	10204851	11	0	11	136	140	222	13	0
2789	42924	26	131	10204853	80	0	80	136	140	222	12	0
2790	42924	26	131	10204853	420	0	420	136	140	222	14	0
2791	42924	26	134	10204853	538	0	538	136	140	222	17	0
2792	42924	26	131	10204864	30	0	30	136	140	222	5	0
2793	42924	26	131	10204864	50	0	50	136	140	222	6	0
2794	42924	26	131	10204921	20	0	20	136	140	222	5	0
2795	42924	26	131	10204921	36	0	36	136	140	222	11	0

These 3 datasets merged together based on the customer ID, account ID fields to create a single final dataset in numeric form as shown in **Table 4.5**. Since the table has more number of columns which cannot fit in on the page here, the snapshot is split into 3 snapshots.

Table 4.5 – Final dataset sample

txn_date	txn_no	txn_account	txn_amount	credit	debit	balance	txn_type	txn_subtype	txn_description	txn_currency	txn_location_type	txn_location_code	target_account	target_country_code	target_bank code	cust_id
43240	392444	10200037	4822	4822	0	4822	25	122	19	136	142	156	0	0	0	20000040
43240	392443	10200037	51	0	51	4771	26	131	8	136	140	222	68106388641	32	250	20000040
43240	392442	10200037	50	0	50	4721	26	133	18	136	140	222	0	0	0	20000040
43235	390913	10200037	63	0	63	4658	26	131	6	136	140	222	56635158	32	575558	20000040
43235	390912	10200037	289	0	289	4369	26	132	3	136	140	222	631747	32	775758	20000040
43235	390911	10200037	2299	0	2299	2071	26	132	2	136	140	222	555823	32	775758	20000040
43234	390612	10200037	28	0	28	2042	26	131	5	136	140	222	587427	32	775758	20000040
43233	390327	10200037	42	0	42	2000	26	131	8	136	140	222	68929298495	32	250	20000040
43233	390326	10200037	439	0	439	1562	26	131	4	136	140	222	367427	32	559059	20000040
43228	388785	10200037	310	0	310	1252	26	131	12	136	140	222	770135	32	555859	20000040
43226	388182	10200037	45	0	45	1207	26	131	8	136	140	222	42256617045	32	250	20000040
43225	387897	10200037	31	0	31	1176	26	132	13	136	140	222	629589	32	555859	20000040
43223	387269	10200037	1848	1848	0	3024	25	122	19	136	142	165	0	0	0	20000040
43223	387268	10200037	53	0	53	2971	26	131	11	136	140	222	265642	32	555705	20000040
43223	387267	10200037	50	0	50	2921	26	133	18	136	140	222	0	0	0	20000040
43221	386858	10200037	290	0	290	2631	26	132	3	136	140	222	631747	32	775758	20000040
43221	386857	10200037	2307	0	2307	324	26	132	2	136	140	222	555823	32	775758	20000040
42912	386840	10200037	1245	1245	15	1554	44	10	15	126	5	88	22920502	32	505959	20000040
42912	386839	10200037	20	0	20	1534	54	96	59	7	13	344	0	0	0	20000040
43217	385941	10200037	7680	7680	0	9214	25	128	1	136	140	222	0	0	0	20000040
43216	385639	10200037	57	0	57	9158	26	131	8	136	140	222	36626266218	32	250	20000040
43212	384445	10200037	93	0	93	9064	26	132	7	136	140	222	848328	32	775758	20000040
43209	383495	10200037	46	0	46	9018	26	131	8	136	140	222	63581834119	32	250	20000040
43207	382902	10200037	297	0	297	8721	26	132	3	136	140	222	631747	32	775758	20000040

cust_id	cust_type	gender	date_of_birth	age	marital_status	residence_country	state	city	postcode	tax_resident_country	birth_country	nationality_country	profession	income_category	kyc_updated_on_date
20000040	4	2	31336	35	2	32	40	86	2079	32	0	0	161	89441	43763
20000040	1	2	31336	35	2	32	40	86	2051	32	0	0	161	130796	43763
20000040	3	2	31336	35	2	32	40	86	2297	32	0	0	161	90848	43763
20000040	2	2	31336	35	2	32	40	86	2170	32	0	0	161	109227	43763
20000040	1	2	31336	35	2	32	40	86	2024	32	0	0	161	99863	43763
20000040	2	2	31336	35	2	32	40	86	2234	32	0	0	161	117057	43763
20000040	1	2	31336	35	2	32	40	86	2365	32	0	0	161	86387	43763
20000040	4	2	31336	35	2	32	40	86	2063	32	0	0	161	153077	43763
20000040	3	2	31336	35	2	32	40	86	2318	32	0	0	161	139278	43763
20000040	1	2	31336	35	2	32	40	86	2107	32	0	0	161	76627	43763
20000040	1	2	31336	35	2	32	40	86	2349	32	0	0	161	122665	43763
20000040	2	2	31336	35	2	32	40	86	2200	32	0	0	161	98709	43763
20000040	2	2	31336	35	2	32	40	86	2286	32	0	0	161	68408	43763
20000040	2	2	31336	35	2	32	40	86	2214	32	0	0	161	82708	43763
20000040	1	2	31336	35	2	32	40	86	2101	32	0	0	161	155096	43763
20000040	1	2	31336	35	2	32	40	86	2052	32	0	0	161	64973	43763
20000040	3	2	31336	35	2	32	40	86	2344	32	0	0	161	158699	43763
20000040	2	2	31336	35	2	32	40	86	2009	32	54	54	161	124185	43763
20000040	2	2	31336	35	2	32	40	86	2009	32	54	54	161	124185	43763
20000040	4	2	31336	35	2	32	40	86	2204	32	0	0	161	117958	43763
20000040	2	2	31336	35	2	32	40	86	2342	32	0	0	161	89233	43763
20000040	1	2	31336	35	2	32	40	86	2081	32	0	0	161	112235	43763
20000040	2	2	31336	35	2	32	40	86	2110	32	0	0	161	113425	43763
20000040	3	2	31336	35	2	32	40	86	2312	32	0	0	161	134196	43763

kyc_updated_on_date	kyc_state	risk_rating	account_no	bsbno	created_on_date	account_type	daily_txn_limit	tfn	statement_delivery_method	issuspicious	Noise
43763	1	0.26014252	10200037	201103	43035	1	2000	827977602	0	0	0
43763	1	0.93122755	10200037	201103	43035	2	10000	827977602	1	0	0
43763	1	0.89822542	10200037	201103	43035	3	5000	827977602	1	0	0
43763	1	0.75943719	10200037	201103	43035	3	2000	827977602	0	0	0
43763	1	0.39710114	10200037	201103	43035	1	5000	827977602	0	0	0
43763	1	0.1324111	10200037	201103	43035	1	10000	827977602	0	0	0
43763	1	0.48519375	10200037	201103	43035	2	2000	827977602	1	0	0
43763	1	0.78414458	10200037	201103	43035	3	20000	827977602	0	0	0
43763	1	0.70074606	10200037	201103	43035	2	20000	827977602	0	0	0
43763	1	0.04232247	10200037	201103	43035	3	2000	827977602	0	0	0
43763	1	0.86309336	10200037	201103	43035	3	5000	827977602	0	0	0
43763	1	0.48797651	10200037	201103	43035	1	2000	827977602	1	0	0
43763	1	0.45038216	10200037	201103	43035	1	10000	827977602	1	0	0
43763	1	0.74269935	10200037	201103	43035	1	10000	827977602	0	0	0
43763	1	0.84397949	10200037	201103	43035	1	2000	827977602	1	0	0
43763	1	0.07453056	10200037	201103	43035	2	5000	827977602	1	0	0
43763	1	0.53981023	10200037	201103	43035	3	5000	827977602	0	0	0
43763	1	0.81029892	10200037	201103	43035	2	20000	827977602	1	0	1
43763	1	0.81029892	10200037	201103	43035	2	20000	827977602	1	1	2
43763	1	0.30525165	10200037	201103	43035	1	2000	827977602	0	0	0
43763	1	0.43723077	10200037	201103	43035	2	20000	827977602	1	0	0
43763	1	0.6835923	10200037	201103	43035	1	5000	827977602	0	0	0
43763	1	0.19122497	10200037	201103	43035	3	20000	827977602	1	0	0
43763	1	0.91581683	10200037	201103	43035	1	20000	827977602	0	0	0

The “IsSuspicious” column in **Table 4.5** presents the labelled classification for training the model. Value 0 represents the transaction as legitimate, value 1 represents the transaction as a suspicious or money laundering transaction. The last column “noise” is added to quickly find the rows which represents noise in the data, this column is removed when the dataset is given as input to training the models.

The generated retail banking customer’s data is more aligned with the developed countries, where the majority of the financial transactions are through card or online, and very few transactions are in the cash. The transaction value is also less considering the higher currency value (most of the transactions are in the range of $10 to $200). Whereas in underdeveloped or developing countries, majority of the financial transactions are in the form of cash and to create the transaction synthetic dataset that aligns with such countries, following factors should be considered:

- proportion of cash transactions should be higher than the digital transactions,
- the transaction values should also be of higher amount (assuming the currency differences), and mostly the withdrawals or deposits are of higher amount,
- the number of transactions with the bank would be less compared transactions of a customer in developed country,
- day to day shopping transactions are in the form of cash,
- transaction thresholds would be different and higher.

Note: For the purpose of this thesis, the transaction data for retail banking customers from Australia is generated and used for modelling.

4.3 Objective 1 - Detect Suspicious Transactions

Objective 1 of this study is to design an effective deep learning model to accurately detect the suspicious money laundering transactions. CNN is considered to address this objective and the following traditional machine learning models are considered to compare the performance of CNN with – RF, SVM and XGBoost.

4.3.1 Suspicious Transaction Detection using CNN

CNN model always goes through multiple rounds of training and testing to check the results by tweaking the hyper parameters. We trained and tested the CNN model several times by changing the hyper parameters that includes epochs, batch size, layers, dropouts, number of units, activation function and finalized the parameters.

Initially, we built a model using bare minimum foundational layers which includes Conv1D, Batch Normalisation, Dropout, Flatten, Dense layer. Then the model was further enhanced by adding and testing with additional layers of Conv1D and Dense layer; changing and testing with dropout percentage; changing and testing the number of units in Conv1D layer; changing the activation function and verifying the model output; changing the number of epochs, batch size and verifying the results.

After trying various combinations of layers and hyper parameters, we settled down on the parameters shown in **Table 4.6** that produced optimum results with 100 epochs and 32 batch size. The model was complied with Adam optimizer (Kingma and Ba 2014) and binary cross entropy loss function (Chollet 2015).

Table 4.6 – CNN architecture hyper parameters

Layer	Parameters
Conv1D	Filters = 32, Kernel size = 2, Input shape = 51980 × 40, Activation = ReLU
Batch Normalisation	Default parameters
Dropout	0.3
Conv1D	Filters = 64, Kernel size = 2, Activation = ReLU
Batch Normalisation	Default parameters
Dropout	0.3
Conv1D	Filters = 128, Kernel size = 2, Activation = ReLU
Batch Normalisation	Default parameters
Dropout	0.3
Flatten	Default parameters
Dropout	0.3
Dense	Units = 512, Activation = ReLU
Dropout	0.3
Dense	Units = 1, Activation = Sigmoid

CNN Prediction Results

For comparing the results of CNN model, several popular traditional machine learning models such as RF, XGBoost and SVM classifiers, were used. All Classifiers considered 70 % data for training (70% data is further split into two parts - 80% for training and 20% for validation) and 30 % data for testing the model. The spilt data is normally distributed with feature column mean at 0 and standard deviation at 1 using standard scaler. This makes it easier to apply the weights and train the model. **Table 4.7** presents the performance metrics of the prediction made by CNN. The table contains the following metrics - confusion matrix, recall **(3.9)**, precision **(3.8)**, accuracy **(3.7)**, AUC score **(3.11)** and $f_\beta$score **(3.10)** with $\beta = 3$.

Table 4.7 – CNN prediction results

CNN			
TN	24515	**FP**	2109
FN	110	**TP**	1114
Recall	91.01	**Precision**	34.56
Accuracy	92.03	**AUC**	98.00
$f_\beta$score	78.23	**T-Time**	70 min

Confusion matrix presents the performance and the errors made by the classifiers for classification. Type-I error is shown by false positives and type-II error is shown by false negatives. Error importance depends on the domain of the classification problem. In case of AML we gave higher importance to type-II error after consulting with AML experts in the industry, type-I error means more alerts and more operational efforts to investigate the alerts, but type-II error indicates the system is not effective in identifying true suspicious transactions which is a bigger risk from compliance point of view as well as allowing the launderers to exploit the financial system for their benefits. Hence, the model is required to produce as less false negatives as possible to improve suspicious transaction detection. Type-II error with high importance implies that recall metric is more important than precision, hence we have considered $f_\beta$score **(3.10)** metric calculation with $\beta = 3$ to reflect higher importance to recall metric.

ROC Curve for CNN

Figure 4.1 shows the Receiver Operator Characteristic (ROC) curve (Pepe 2000) plotted to show the binary classification capability of CNN classifier. The curve is formed by plotting the true positive rate against false positive rate. The true positive rate indicates the number of scenarios that are correctly predicted as positive out of all positive scenarios as presented in equation **(4.1)**.

$$true\ positive\ rate\ = \frac{TP}{TP + FN} \tag{4.1}$$

The false positive rate indicates the number of scenarios that are incorrectly predicted as positive out of all negative scenarios as presented in equation **(4.2)**.

$$false\ positive\ rate\ = \frac{FP}{TN + FP} \tag{4.2}$$

ROC curve measures the ability of the classifier in distinguishing the transactions as suspicious or legitimate. Higher the curve, better the model's ability to distinguish between the classification category. Closer the curve to the upper left corner, better the performance or accurate the predictions are and the closer the curve to the diagonal line less accurate the predictions are.

One of the key characteristics of the ROC curve is, it doesn't depend upon the class distribution, and hence it shows the model performance very clearly. In this study the data is highly imbalanced, hence ROC curve is quite helpful to show the prediction performances of the models. To represent the performance of a model using a single metric, an area under the ROC curve is calculated which is normally called as AUC.

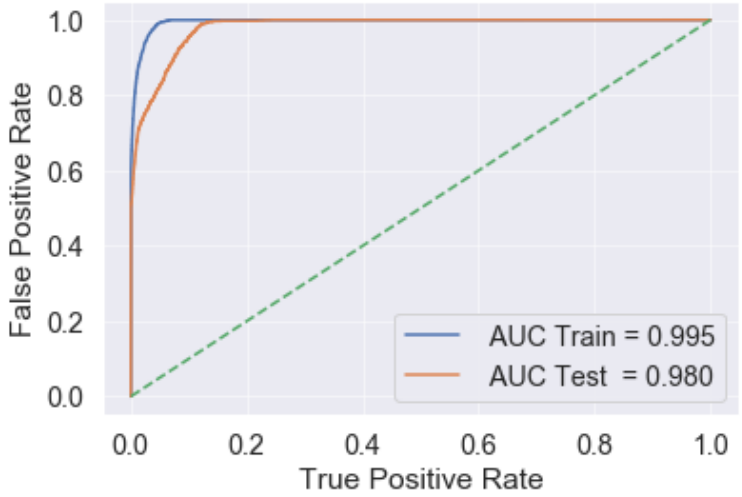


Figure 4.1 – ROC curve for CNN Model

4.3.2 Validation of CNN Results using RF

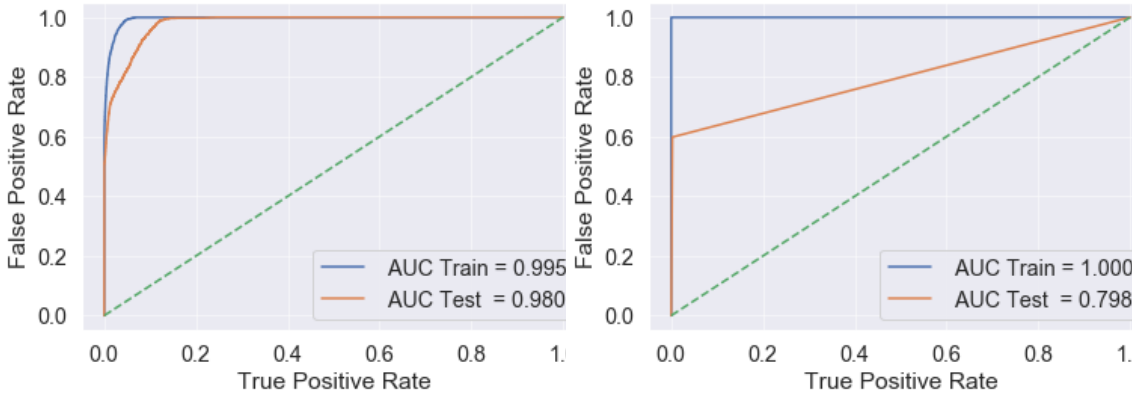
To study how other classifiers perform with the synthetic data and compare the performance with CNN, we trained and tested the RF model. **Table 4.8** presents the performance metrics for CNN and RF classifier.

Table 4.8 – CNN results comparison with RF results

CNN				RF			
TN	24515	**FP**	2109	**TN**	26532	**FP**	69
FN	110	**TP**	1114	**FN**	501	**TP**	746
Recall	91.01	**Precision**	34.56	**Recall**	59.82	**Precision**	91.53
Accuracy	92.03	**AUC**	98.00	**Accuracy**	97.95	**AUC**	79.80
$\mathbf{f_\beta score}$	78.23	**T-Time**	70 min	$\mathbf{f_\beta score}$	61.97	**T-Time**	6 sec

The comparison study between CNN and RF shows the following key observations (read in conjunctions with **Table 4.8**):

- CNN is able to manage the compliance risk better than RF considering the fact that $f_\beta$score **(3.10)** of CNN is greater than RF and false negatives of CNN are less than RF.
- RF is able to manage the operational efforts of AML investigation team way better than CNN since the false positives produced by RF are negligible compared to CNN. It is important to note here, saving on operational efforts at the cost of increasing the compliance risk is not a practical approach considered by financial institutions.
- For the same dataset, the training time of RF is very negligible (7 seconds) compared to CNN (1 hr 10 minutes). However, it is important to note that the key limitation of the RF classifier is that the more number of trees can be make the classifier to too slow in spite of parallelism used by classifier to use the available resources (CPU and memory) effectively.



a. CNN **b. RF**

Figure 4.2 – ROC curve – (a) CNN and (b) RF

ROC Curve for CNN and RF

Figure 4.2 shows the ROC curve for CNN and RF. The AUC curve for CNN is shows better results than RF.

Feature importance plot for RF

Machine learning classifiers predict the outcome by giving weightage to certain features and that information helps to interpret the decisions made by the classifier. Feature importance score is one such method that is used to construct a graph of features against the importance score which shows the most important features considered by classifier while making the predictions. The feature importance plot helps in many ways, such as correcting the dataset if the importance is given to irrelevant features as per understanding of domain to demonstrate high interpretability of the model. **Figure 4.3** shows the feature importance plot for the RF classifier, which shows that the model has given the high importance to transaction sub type, transaction description, transaction amount, transaction location type, and debit features while making the predictions.

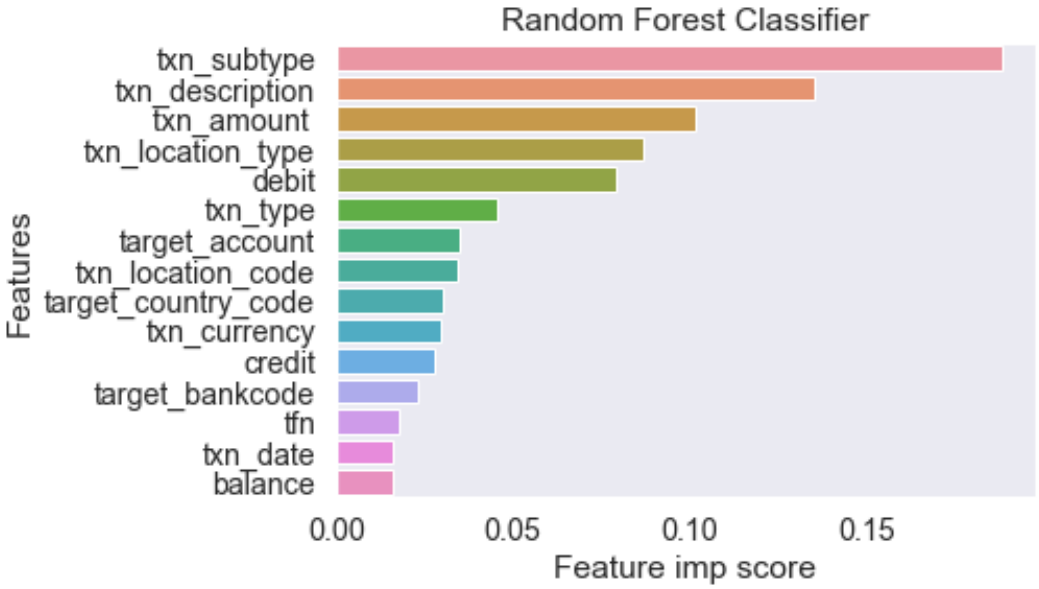


Figure 4.3 – Global interpretation of RF classifier

4.3.3 Validation of CNN Results using SVM

SVM is one of the most frequently used supervised machine learning model used for classification and regression analysis that produces significant accurate predictions with less computation power. To study how the SVM classifier performs with the synthetic data and compare the performance with CNN, we trained and tested SVM classifier. **Table 4.9** presents the performance metrics for CNN and SVM.

Table 4.9 – CNN results comparison with SVM results

CNN				SVM			
TN	24515	**FP**	2109	**TN**	26426	**FP**	175
FN	110	**TP**	1114	**FN**	884	**TP**	363
Recall	91.01	**Precision**	34.56	**Recall**	29.10	**Precision**	67.47
Accuracy	92.03	**AUC**	98.00	**Accuracy**	96.20	**AUC**	83.60
$f_\beta$score	78.23	**T-Time**	70 min	**$f_\beta$score**	30.86	**T-Time**	4.4 min

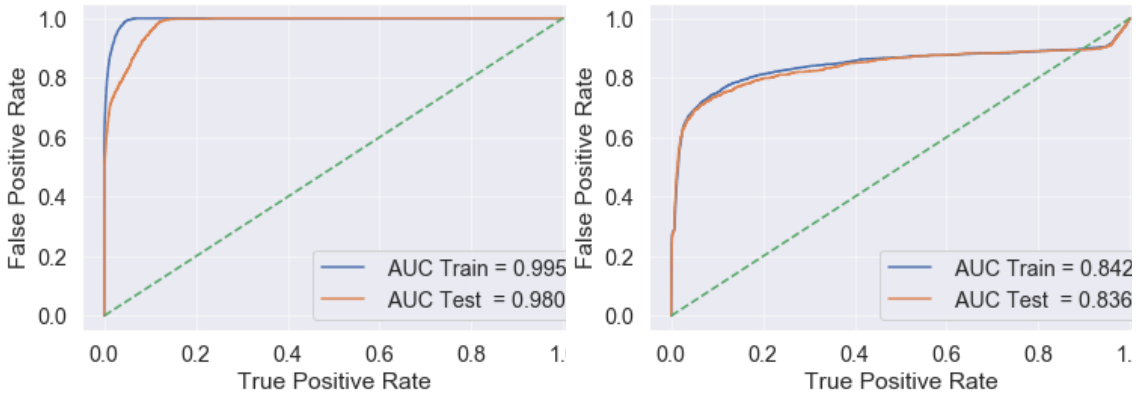
The comparison study between CNN and SVM shows the following key observations (read in conjunction with **Table 4.9**):

- CNN is able to manage the compliance risk better than SVM considering f_{β}score **(3.10)** of CNN is greater than SVM and FNs of CNN are less than SVM.
- SVM is able to manage the operational efforts of AML investigation team way better than CNN since the false positives produced by SVM are very less compared to CNN. It is important to note here, saving on operational efforts at the cost of increasing the compliance risk is not a practical approach considered by financial institutions.
- For the same dataset, the training time of SVM is very less (4.4 minutes) compared to CNN (1 hr 10 minutes).

The limitations of SVM are, it doesn't work well with a large dataset and when the dataset has overlapping classes. In the case of the finance domain, there are millions of transactions per day and there are several overlapping transactions in the dataset. Overlapping transaction are that transactions which appears similar to legitimate transactions but actually they are suspicious money laundering transactions.

ROC Curve for CNN and SVM

Figure 4.4 shows the ROC curve for CNN and SVM. The AUC curve for CNN is shows better results than SVM.



a. CNN **b. SVM**

Figure 4.4 – RoC curve – (a) CNN and (b) SVM

Feature importance score plot for SVM

Figure 4.5 shows the feature importance plot for the SVM. It is important to note here that the SVM classifier has given a high importance score to different features than RF. According to feature importance plot the KYC state, target account, transaction date are contributing higher in predicting the suspicious transactions.

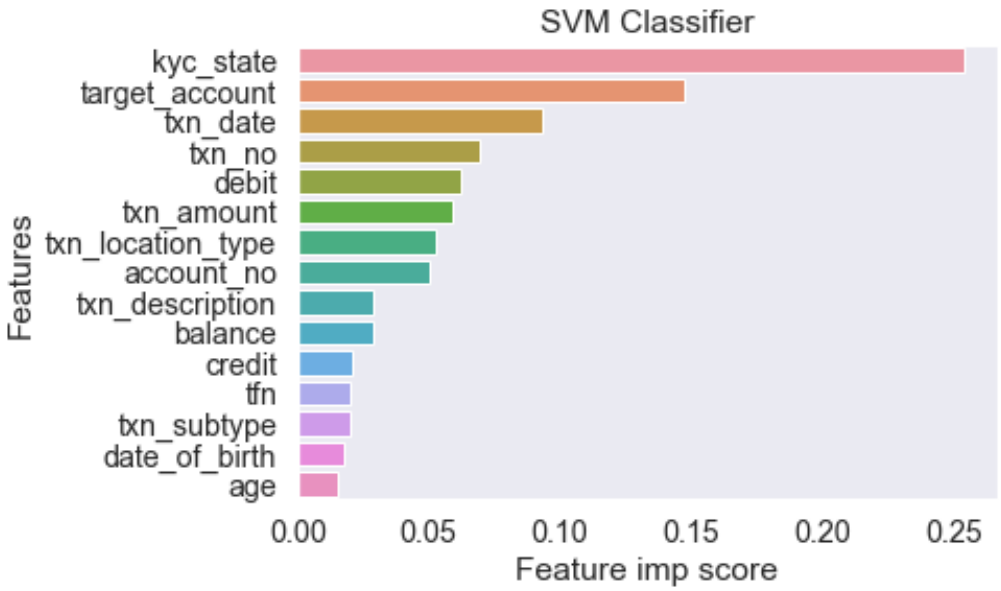


Figure 4.5 – Global interpretation of SVM classifier

4.3.4 Validation of CNN Results using XGBoost

XGBoost algorithm is a supervised machine learning classifier based on the decision tree used for classification and regression analysis that produces significant accurate predictions with less computation power. To study, how the XGBoost classifier performs with the synthetic data and compares the performance with CNN, we trained and tested the XGBoost classifier. **Table 4.10** presents the performance metrics for CNN and XGBoost classifier.

Table 4.10 – CNN results comparison with XGBoost results

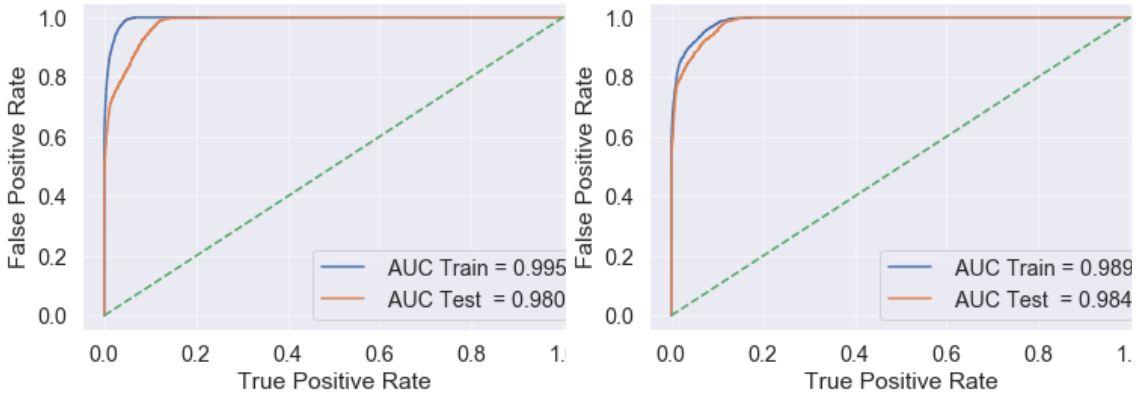
CNN				XGBoost			
TN	24515	**FP**	2109	**TN**	26496	**FP**	105
FN	110	**TP**	1114	**FN**	497	**TP**	750
Recall	91.01	**Precision**	34.56	**Recall**	60.14	**Precision**	87.72
Accuracy	92.03	**AUC**	98.00	**Accuracy**	97.84	**AUC**	98.40
$\mathbf{f_\beta score}$	78.23	**T-Time**	70 min	$\mathbf{f_\beta score}$	62.09	**T-Time**	16 sec

The comparison study between CNN and XGBoost shows the following key observations (read in conjunction with **Table 4.9**):

- CNN is able to manage the compliance risk better than XGBoost considering the fact that $f_\beta score$ **(3.10)** of CNN is greater than XGBoost and false negatives of CNN are less than XGBoost.
- XGBoost is able to manage the operational efforts of AML investigation team way better than CNN since the false positives produced by XGBoost are very less compared to CNN. It is important to note here, saving on operational efforts at the cost of increasing the compliance risk is not a practical approach considered by financial institutions.
- For the same dataset, the training time of the XGBoost is very negligible (16 seconds) compared to CNN (1 hr 10 minutes).

Few key advantages of XGBoost classifier are, it can leverage the parallel processing ability of computer, able to take care of missing data well. The limitations are, XGBoost is very sensitive to outliers and doesn't perform so well when the training samples are way less. In the case of the finance domain, there are millions of transactions per day and there is always a scarcity of labelled training data.

ROC Curve for CNN and XGBoost



a. CNN **b. XGBoost**

Figure 4.6 – ROC curve – (a) CNN and (b) XGBoost

Figure 4.6 shows the ROC curve for CNN and XGBoost method. The AUC test curve for XGBoost shows better results than CNN method.

Feature importance score plot for XGBoost

Figure 4.7 presents the feature importance plot for the XGBoost which shows the highly contributing features for predictions as transaction subtype, KYC state, tax file number (TFN) and transaction location type.

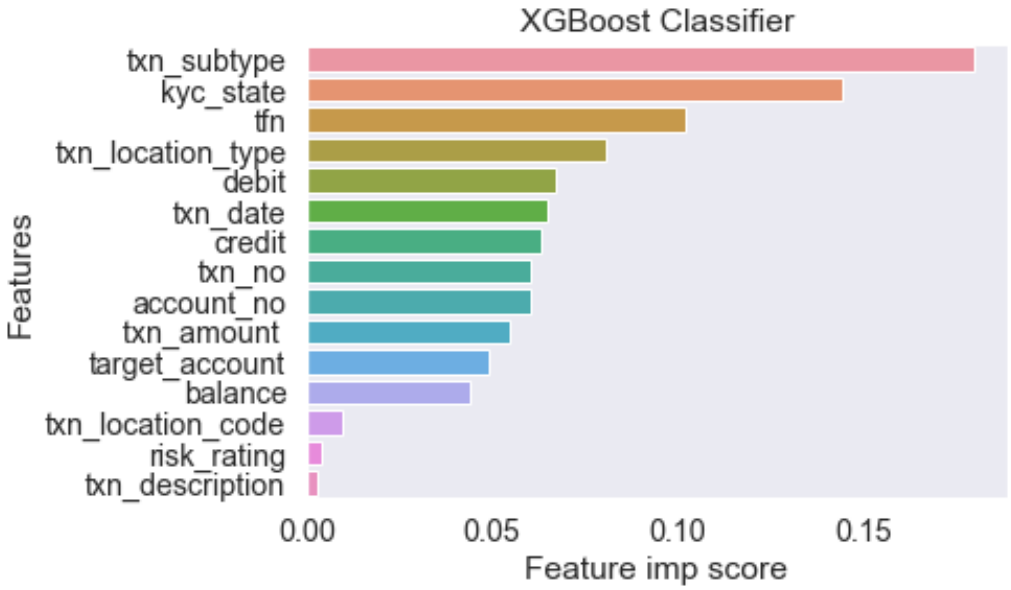


Figure 4.7 – Global interpretation of XGBoost classifier

4.3.5 Discussion

Banks are using rule-based AML systems for generating the alerts based on the rules specified by regulatory (Finance 2018). This includes cash transactions, international fund transfers and suspicious money laundering transactions. Suspicious transaction alerts are raised by the system based on the rules defined as per money laundering patterns detected earlier. It is a difficult task to keep these rules up to date considering a pace at which number of transactions are growing and new fraud patterns are innovated by fraudsters (Jullum et al. 2020). This causes more false positive alerts by system and impacts the operational efficiency. AML officers are expected to investigate each alert thoroughly before reporting to the regulatory. AI/ML based systems that churns out the huge data and identify the suspicious transactions acts as decision support system for AML officers. These systems are trained using the historical money laundering patterns which includes alerts generated by system, details captured in case report and SMR. The key goal of AI/ML based AML systems is to help the AML officer by identifying suspicious transactions along with adequate evidences (Finance 2018). Non-identified suspicious transactions poses a serious risk to a Bank from compliance point of view which leads to heavy penalties (Office of Parliamentary Counsel 2019), (Network 2021). Hence, financial institutions focus more on ensuring that they do not miss out on any suspicious transactions, even though they have to investigate more alerts. The core characteristics of deep learning to identify features automatically and detecting the suspicious transactions with high accuracy can help to identify suspicious transactions by minimising the false positive and false negatives.

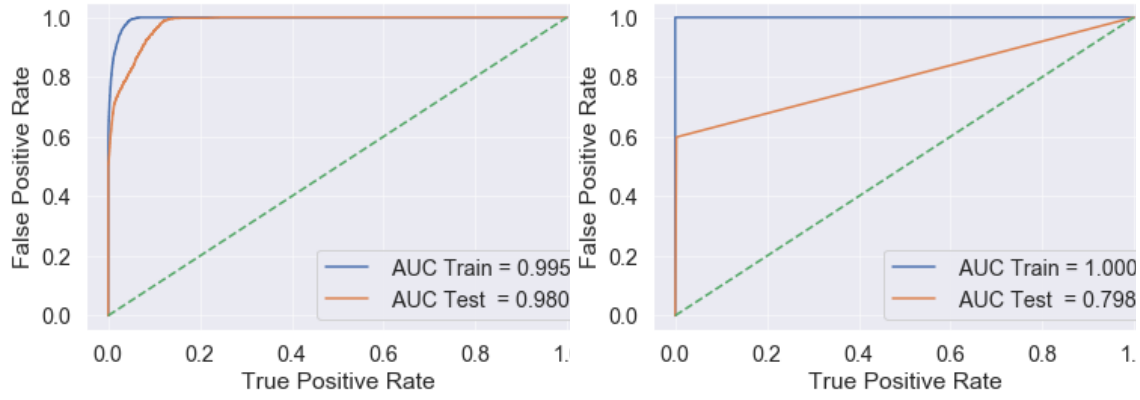
We have generated a synthetic data containing customers, accounts and transactions; and used the same for research. In reality, the data used for money laundering detection is highly imbalanced, meaning the legitimate transactions are high in number and suspicious transactions way less compared legitimate, hence we followed a similar pattern. Our dataset proportion was 95.64 % to 4.36 % for legitimate and suspicious transactions, respectively. Since the data is highly imbalanced, accuracy **(3.7)** was not considered as a primary metric. We considered f_{β}score **(3.10)** score as a key metric of performance measurement with $\beta = 3$ to value recall more than precision.

Table 4.11 – Prediction Results CNN, RF, XGBoost and SVM

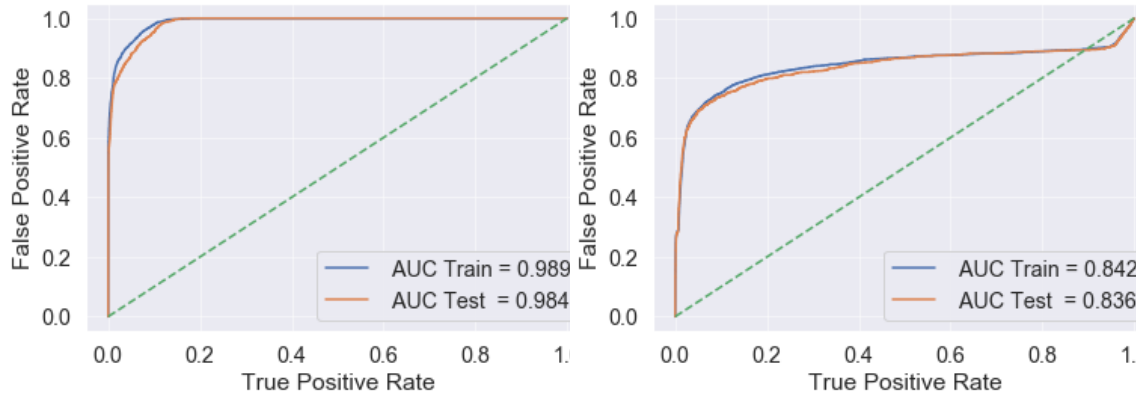
CNN				RF			
TN	24515	**FP**	2109	**TN**	26532	**FP**	69
FN	110	**TP**	1114	**FN**	501	**TP**	746
Recall	91.01	**Precision**	34.56	**Recall**	59.82	**Precision**	91.53
Accuracy	92.03	**AUC**	98.00	**Accuracy**	97.95	**AUC**	79.80
$\mathbf{f_\beta score}$	78.23	**T-Time**	70 min	$\mathbf{f_\beta score}$	61.97	**T-Time**	6 sec
XGBoost				**SVM**			
TN	26496	**FP**	105	**TN**	26426	**FP**	175
FN	497	**TP**	750	**FN**	884	**TP**	363
Recall	60.14	**Precision**	87.72	**Recall**	29.10	**Precision**	67.47
Accuracy	97.84	**AUC**	98.40	**Accuracy**	96.20	**AUC**	83.60
$\mathbf{f_\beta score}$	62.09	**T-Time**	16 sec	$\mathbf{f_\beta score}$	30.86	**T-Time**	4.4 min

This study has successfully trained and tested the Conv1D based CNN model for classifying the transaction as suspicious and legitimate. The CNN results are compared with RF, XGBoost and SVM classifier on the same dataset. **Table 4.11** shows the results of CNN, RF, SVM and XGBoost.

As per the $f_\beta$score, we found that CNN model outperforms RF, XGBoost and SVM model, which means the false negatives are less compared to other models and a good indication from risk reduction point of view. However, the false positive number of CNN model is higher than other models, which means more operational efforts for investigation. A view of AML subject matter experts is the first priority is to reduce the risk, ensure regulatory compliance and then reduce the operational efforts. Deep learning methods are believed that they perform better with large data set, and in AML case the volume of transactions are growing day by day with the digital banking which makes deep learning an apt method to apply.



a. CNN **b. RF**



c. XGBoost **d. SVM**

Figure 4.8 – ROC curve – (a) CNN, (b) RF, (c) XGBoost and (d) SVM

Figure 4.8 shows the CNN ROC curve comparison with machine learning classifiers RF, XGBoost and SVM.

The CNN model is used effectively to identify the highest number of suspicious transactions with high $f_\beta$score (78.23) and recall rate (91.01). The synthetic dataset contains the features that are very similar to the dataset features of AML systems in bank such as customer, accounts and transactions. The results also found some customer profile attributes as key contributors in detecting suspicious transactions.

The results show CNN model outperforms other model considering FN identified, Recall and $f_\beta$score. The key characteristics of deep learning models is, the accuracy improves

with the increase in data size, and in case of finance domain, millions of transactions are generated every day. That gives a good indication for the application of the CNN model in finance domain.

4.3.5.1 Limitations

The limitation of this research is, the deep learning and machine learning models are trained and tested on synthetic data and not on real finance data from banks. All the experiments are performed on a synthetic data which has limited type of customers, products (business loan, insurance etc) owned by customers and type of transactions, whereas in Banks there would be much more complex transaction dataset and the access to external data sources such as PEP screening, world-check list and social network datasets.

The Banks have different types of products and associated transactions such as accounts, loans, credit cards, insurance, securities, mobile transfers, cheques, money drafts; different types of customers including retail customers, small businesses, business banking, institutional banking, investment banking; regulations based on the counties in which the Bank operates. Hence, it is important to test the model holistically by considering all such scenarios and data in a stage manner. The model is tested on a very limited set of data, whereas in real banking scenario there would be millions of transactions per day, hence the model's performance considering the continuous flood of transaction must be tested. Author believes that the suspicious transaction detection heavily depends upon the availability of historical money laundering transactions data for training. To qualify a transaction(s) as money laundering, it goes through rigorous process which includes multi-stage review (include alert, case, SMR and criminal investigation) and creation of different reports containing additional evidences. Hence, the quality of consolidated suspicious transaction scenario data would also contribute in the identifying suspicious transactions by models.

4.3.5.2 Future work

At core, deep learning models are designed to work with image type of data, hence the recommendation for future work would be to covert the tabular dataset into images and apply Conv2D CNN model to check the outcome for suspicious transaction detection.

Most of the XAI techniques such as SHAP (Lundberg and Lee 2017), LIME (Ribeiro et al. 2016) supports image data for explaining the predictions made by deep learning models, hence converting the tabular data into image and then predicting suspicious transactions using Conv2D CNN models can open up several XAI options to interpret the predictions.

4.4 Objective 2 - Explain Predictions

Objective 2 of this thesis is to design a method to explain the reasoning behind the predictions made by the chosen deep learning method which is CNN. To address this objective, the SHAP method is chosen and the same is applied on CNN to explain the predictions made by CNN classifier.

4.4.1 Explain CNN Predictions using SHAP

Models are interpreted at local and global level. Local interpretation focuses on bringing out the reasoning behind the individual prediction, whereas global interpretation focuses on how model behaves in general. This section describes the local interpretation using SHAP. After training and testing the CNN classifier, we applied SHAP to interpret the feature importance considered by CNN while making the predictions. A DeepExplainer class from the SHAP library is used to generate the shap values for a test data of 24410 rows and 40 columns, which took approximately 20 minutes to generate the shap values for each data value in the test dataset. The availability of shap values for each data element on a record, makes it possible to explain the individual record prediction by applying these shap values to provide a view of highly contributing features leading to a decision.

Figure 4.9 and **Figure 4.10** show the force plot representing the interpretation of individual record prediction made by a CNN classifier. The base value shown on the plot is the average value of the target variable across the dataset we passed to DeepExplainer class. Each arrow strip shows the impact of its associated feature on pushing the target variable away or close to base value. Red strips show that their associated feature pushes the value on the higher side (indicating transaction being suspicious) with respect to base value, whereas the blue strips show that the associated feature pushes the value on the lower side (indicating transaction being legitimate).

Figure 4.9 presents the interpretation of one record that was predicted as a suspicious transaction by CNN classifier. The key features that are contributing to the prediction are credit, txn_amount, txn_description, kyc_state and txn_currency. The width of the red or blue strips indicates the positive or negative impact made by the features associated with the strips respectively. Here positive impact pushes the score towards 1 (suspicious transaction is denoted by 1), negative impact pushes the score towards 0 (legitimate transaction is denoted by 0). Features with higher impact are located closer to the boundary that separates blue and red strips. Higher the impact, bigger the size of strip.

Table 4.12 shows the details of an original transaction record that is predicted as a suspicious transaction. The inference from force plot and original transaction record is, the customer associated with the transaction account appears to be a student, transaction is done in cash with decent large amount, the cash is deposited into account from ATM deposit machine, and being a student the account balance appears to be high. They are showing a positive correlation to towards the transaction being suspicious. Whereas the kyc_state shows a negative correlation with the prediction. This inference makes the transaction as suspicious and worth investing further.

Please note, the suspicious transactions identified by CNN classifier are as per the training data provided. In reality, the suspicious transaction goes through a multiple level of investigation (by Bank, regulatory, investigation agency to court) before it is qualified as a true money laundering transaction. The real value addition by shap explanation for investigation is the identification of key contributing features for a decision.

Figure 4.10 presents the interpretation of a record that was predicted as a legitimate transaction by CNN classifier. The key features that are contributing to the prediction are – txn_description, txn_location_code, txn_location_type, txn_currency, txn_subtype.

Table 4.12 shows the details of an original transaction record that is predicted as a legitimate transaction. The inference from force plot and original transaction record is, the associated customer with the transaction is a married female having an age of 55, and the identified transaction is auto-debit transaction that is paid online for health insurance in a local currency of her birth country – which concludes the transaction is legitimate. In reality, the legitimate transactions will not be screened, however, in this paper we are

looking at the interpretation to validate the reasoning presented by SHAP for the prediction made by CNN.

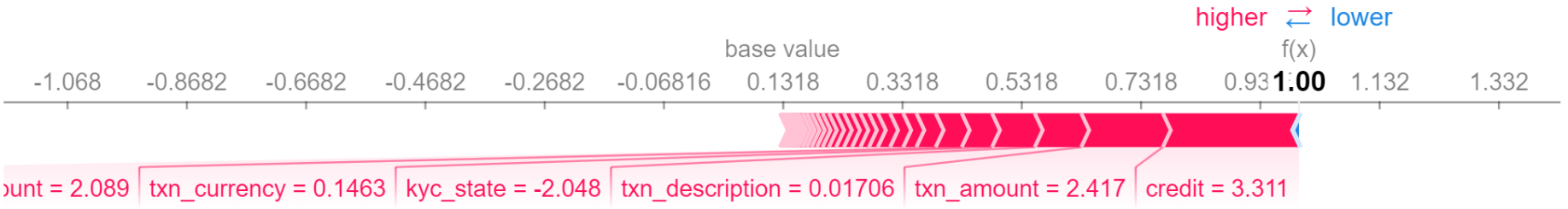


Figure 4.9 – SHAP force plot for suspicious transaction

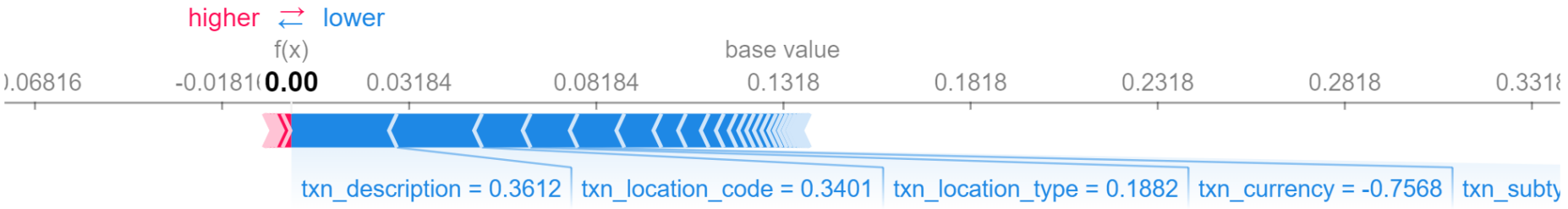


Figure 4.10 – SHAP force plot for legitimate transaction

Table 4.12 – Original transaction records associated with CNN predictions

Features	**Suspicious Txn**	**Legitimate Txn**
txn_date	7/12/2017	6/02/2018
txn_no	339549	359932
txn_account	10300015	10202449
txn_amount	**6,000.00**	**322.00**
credit	**6,000.00**	-
debit	-	**322.00**
balance	**52,659.00**	16,054.00
txn_type	Credit	Debit
txn_subtype	Cash deposit	**Auto-debit**
txn_description	**Cash deposit**	**Health insurance**
txn_currency	**AUD**	**AUD**
txn_location_type	**ATM**	**Online**
txn_location_code	448	**222**
target_account	0	891141
target_country_code	0	Australia
target_bankcode	0	559059
cust_id	20000736	20002452
cust_type	**Student**	Individual
gender	Male	**Female**
date_of_birth	24/09/1992	26/05/1965
age	28	**55**
marital_status	Single	**Married**
residence_country	Australia	Australia
state	New South Wales	New South Wales
city	Sydney	New Castle
postcode	2358	2361
tax_resident_country	Australia	Australia
birth_country	Overseas country	Australia
nationality_country	Overseas country	Australia
profession	Student	Labourers

Features	Suspicious Txn	Legitimate Txn
income_category	4000	77668
kyc_updated_on_date	**22/04/2017**	13/09/2019
kyc_state	**Active**	Active
risk_rating	0	0.463290428
account_no	10300015	10202449
bsbno	203901	201807
created_on_date	22/04/2017	23/08/2017
account_type	Savings	Savings
daily_txn_limit	3000	2000
tfn	999528645	968305061
statement_delivery_method	Not set	Online

The true advantage of the SHAP technique is an explainability of each prediction at the individual record level that helps the user to take a quick look at the original record and the contributing features in the prediction. Author applied state-of-the-art SHAP XAI method on CNN classifier to interpret the decisions made by CNN at individual transaction level. Interpretation shown by SHAP has identified satisfactory reasoning for the predictions made by CNN, and by inspecting those indicators along with the original records containing customer, account and transaction details gives a good insight to AML officer to decide if the identified suspicious transaction is a false positive or indeed worth investigating further.

Figure 4.3, Figure 4.5, Figure 4.7 show the global interpretation or the feature importance graphs for RF, SVM and XGBoost classifiers respectively which are used to compare with CNN. Each feature contributes differently in all 3 models. These are also opaque models and SHAP has the capability to explain the individual predictions, however, that that is not an objective of this study.

4.4.2 Explain CNN Predictions using LIME

As described in section **3.4.1,** Conv1D CNN is chosen to detect suspicious transactions in this study. We intended to use LIME on Conv1D CNN to explain the predictions, however, it was found that LIME does not support Conv1D CNN classifier. LIME

supports Conv2D CNN classifier, which requires the input data in 3 dimensional format. The tabular data which is a two dimensional data, should be converted to three dimension (or image) to make it consumable for Conv2D CNN model. Since converting the tabular data into image data was a separate topic of research, the same was not considered in the scope of this thesis.

4.5 Summary

By keeping both the research objectives in mind, the study has successfully conducted the experiments of training and testing the models on created synthetic finance data to meet the objectives.

As per objective 1, the 1D CNN model is trained and tested to detect the suspicious transactions. After training and testing the RF, SVM and XGBoost models for detecting suspicious transactions; we compared the performance of CNN with these models. The comparison result showed that CNN model outperformed other models in detecting most true positive suspicious transactions. CNN model has shown the lowest false negative predictions which indicates the model is able to catch maximum injected suspicious transactions compared to RF, SVM and XGBoost. This is an excellent result from a financial institutions point of view, because it reduces the risk of skipping the true money laundering transactions and reduce the non-compliance risk. CNN model outperformed RF, SVM and XGBoost with respect to f_{β}score. CNN model has shown a high number of false positive predictions that indicates more operational efforts by AML investigators to scan through the alerts and hence more operational cost. However, from a practical point of view, reducing compliance risk is far more important than focusing on operational cost.

As per objective 2, the study has demonstrated that SHAP XAI technique can be used effectively on CNN model that brings out the feature importance at the individual record level for the predictions. SHAPs ability to show the interpretation of CNN predictions by pinpointing the contribution score of each feature to reach a decision at individual record level would help the AML officer investigate the suspicious transactions quickly and

improve the effectiveness of AML controls at Banks and enable the adoption of deep learning in AML domain.

5 CONCLUSIONS AND FUTURE WORK RECOMMENDATIONS

5.1 General Conclusion

This study aimed to predict the suspicious money laundering transactions using deep learning technique and design a method to explain the predictions made by chosen deep learning method.

This research required the finance transaction data for modelling and experiments. After interacting with few financial institutions, it was understood that the finance data is very sensitive and private in nature hence it wasn't possible to obtain the same for research purpose. Hence, a synthetic data generator software was written to generate the data automatically by defining and implementing the rules for creating customers, accounts and transactions. The synthetic data generated from the software was used for modelling and experiments performed in this research.

To address the first objective of the research which is "detect suspicious money laundering transactions", a novel 1D CNN model is implemented for suspicious transaction detection for the first time as of our best knowledge. For appraising the performance of the CNN classifier, three traditional machine learning classifiers (RF, SVM and XGBoost) are implemented and trained on the same data as of CNN.

To address the second objective of the research which is "explain the predictions made by chosen deep learning method", a SHAP XAI technique is designed and applied on CNN classifier to explain the predictions made by CNN for detecting suspicious transactions which is also a first study in AML domain.

In suspicious transaction classification, false positive indicates more operational efforts and false negative indicates high risk of non-compliance; considering a non-compliance risk over operational efforts, a recall metric is given higher weightage over precision and along with the confusion matrix, recall, precision and accuracy, $f_\beta$score with $\beta = 3$ used for measuring the performance of models.

5.2 Conclusion of Objective 1

Objective 1 of this thesis was to design an effective deep learning model to accurately detect the suspicious money laundering transactions, which is addressed by implementing a 1D CNN model. The conclusion of the study are:

Pros:

- CNN model successfully trained and tested using synthetic tabular data.
- CNN model successfully identified the synthetically injected money laundering transactions far better than other machine learning models (RF, SVM and XGBoost) that are used for comparison.
- CNN model has shown the lowest (110) false negative predictions which indicates the model is able to detect maximum injected suspicious transactions compared to RF that produced 501 false negatives, SVM produced 884 and XGBoost produced 497. Low false negative is an excellent result from an organisation point of view, because it reduces the risk of skipping the true money laundering transactions and reduce the AML compliance risk.
- CNN model outperformed RF, SVM and XGBoost with respect to $f_\beta$score with beta = 3. CNN $f_\beta$score is the highest 78.23 followed by XGBoost score as 62.09, RF score as 61.97 and SVM score as 30.86.

Cons:

- CNN model has shown high number of false positive predictions (2109) as compared to RF that produced 69, SVM produced 175 and XGBoost produced 105. High false positive result indicates more operational efforts by AML investigators to scan through the alerts and hence more operational cost.
- CNN model took highest training time – 70 minutes, whereas the lowest training time was taken by RF as 6 seconds, followed by XGBoost which took 16 seconds, and SVM took 4.4 minutes.

It is important to note that financial institutions or banks value compliance risk much higher than the operational costs involved in investigating the alerts, hence in the business context 1D CNN model is a best choice of classifier for suspicious transaction detection.

5.3 Conclusion of Objective 2

Objective 2 of this thesis was to design a method to explain the reasoning behind the predictions made by the chosen deep learning method. To address this objective, the research implemented 1D CNN deep learning classifier to detect suspicious money laundering transactions and explained the CNN predictions by implementing and applying SHAP XAI technique over CNN model. The conclusion of the study are:

- SHAP was applied successfully on CNN to determine the positive or negative contribution (using Shapley values) levels of each feature value on the prediction made by 1D CNN.
- Original transaction records along with the SHAP force plots were analysed manually and found that it was possible to comprehend the rationale behind the predictions made by CNN.
- The SHAP results were validated with the AML subject matter expert from the industry, who acknowledged that the SHAP force plot along with relevant original record can assist in investigating the alerts.
- Hence SHAP method can be considered as one of the important solution component while designing the decision support systems for AML investigators at Banks and FIUs.
- The SHAP method took around 20 minutes to generate the Shapley values for 24410 records, which can be a concern considering millions of records in real banking scenario.

5.4 Practical Implications

Banks establish several AML controls to combat money laundering such as KYC, CDD, risk profiling, watchlist screening, sanctions, employee training programs, and transaction monitoring for suspicious activities. Considering the continuous increase in transactions, changing fraud patterns and regulatory landscape, any method that is able to help identify the suspicious transactions is of a great value to a bank. Primarily, the rule-based AML systems (usually bought from established software product companies who specialise in finance crime) are used by banks for transaction monitoring which has its

pros and cons. However, the rule-based system is an essential and important system when it comes to detect the transactions for reporting as per the thresholds defined by regulatory (example – IFTIs). From AML perspective, pros are, rule-based systems are configured as per the regulatory rules and historically identified money laundering scenarios. Cons are, practically it is hard (or nearly impossible) to keep the system up to date to ensure rules and money laundering scenarios are updated; this is primarily due to constantly changing regulatory landscape and money laundering patterns. Hence banks are experimenting to use AI/ML technology to assist them in detecting the suspicious transactions and reduce the compliance risk. Having said that, rule-based systems are still essential especially to detect the transactions based on the regulatory rules.

The outcome of this research, a novel method of detecting suspicious transactions using CNN method would be beneficial for a Banks especially in the developed countries (because the models are trained and tested on the transaction dataset containing majority number of transactions in the form or card or online). The recommendation for using this model would be – first ensure the data attributes and data is in the required format, normalise the data, ingest the actual money laundering transactions as per the historical records available in Banks as a labelled training data; train and test the model. It is important to note that, identification of suspicious transactions using CNN would be one of the component in the "to-be" AI-AML system. The CNN method should be used together with other components to further improve the effectiveness of the outcome, that includes customer segmentation, risk profiling, social network analysis, pep screening, sanctions screening etc. to establish adequate ground to define suspiciousness of the transactions and report suspicious matter report for compliance. Author believes that CNN method would help improve the effectiveness of detecting suspicious money laundering transactions leading to reduction of compliance risk and operational cost required for screening and investigation of the false positive alerts (Finance 2018), (Al-Shabandar et al. 2019).

Further the "to-be" AI-AML system of a bank can be extended by using a novel method of "explaining the CNN predictions using SHAP XAI framework" implemented in this research. The explanations generated by post-hoc SHAP methods for CNN predictions would help AML officers to view the reasoning behind the predictions and gather the

relevant evidences to enhance the investigation, and report the suspicious transactions with sufficient evidences and high confidence. Explainable AML system would help building the trust (PricewaterhouseCoopers 2017) and drive the adoption of AI/ML technologies in AML domain which can help to manage the compliance risk, keep the brand reputation intact and avoid hefty penalties for non-compliance.

5.5 Research Drawbacks and Limitations

The limitation of this research is, the models are not trained and tested on real transaction data. All the experiments are performed on a synthetic data which has limited type of customers, products owned by customers and type of transactions, whereas in Banks there would be much more complex transaction dataset and the access to external data sources such as PEP screening, world-check list and social network datasets.

5.6 Recommendations for Future Work

Money laundering crime is detected by identifying suspicious transactions, money laundering patterns from the transactions, launderer's gangs by the flow of money. This requires huge manual effort to investigate transactions and now considering the growth in the transactions per day, this is just beyond imagination to think it as a manual task. This challenge can be very well addressed by the data analytics, AI/ML type of technological solutions. However, the application of this technology is still in the inception phase in the industry. One of the key reason is, the research in AML domain is limited primarily due to unavailability of the real finance data, lack of trust or confidence in the decisions made by machine learning models, and the non-explainability of the decisions made by non-interpretable models. Hence following are the recommendations for the future work:

1. **Test with real finance data** – This research has successfully applied deep learning CNN classifier to detect suspicious money laundering transactions using synthetic data and further explained the CNN predictions using SHAP method. However, this is not tested on the real transaction data from the bank, hence the future recommendation is to train and test the models designed in this thesis on real transaction data from a bank.

2. **Transactions for developing countries –** the thesis has generated the banking transaction data that aligns more with the developed countries where cash transactions are minimal. It is recommended to generate the transaction dataset for the developing countries where cash transactions are more, and of higher amounts considering the currency value; and test the model on the same.

3. **Research collaboration** - Research in AML should be a collaborative work of financial institution and research community / academicians to ensure better results. If collaboration is not feasible then ensure near-real time synthetic transaction data is generated to conduct experiments.

4. **Deep learning techniques** - Deep learning models are designed to work with image type of data, hence the recommendation for future work would be to convert the tabular dataset into images and apply Conv2D CNN model to check the outcome for suspicious transaction detection. This would also open up several options of XAI techniques that supports image data such as LIME.

5. **XAI techniques** – It is a need of AML domain to have an explanation of the decisions made by non-interpretable models. As per the literature, there are several XAI techniques researched so far, however, the application of the same on CNN classifier in AML domain does not exist. Hence, the recommendation is to explore state-of-the-art XAI techniques that can help explain / interpret the decisions made by non-interpretable models.

6. **Reinforcement learning** - One of the practical aspect of AML domain is, each suspicious transaction alert goes through a manual review and investigation process before it is reported to the regulatory; hence the recommendation is to use the reinforcement learning methods and leverage the alert review / investigation outcome knowledge as a labelled training data to improve the performance of the model continuously.

7. **Graph mining and social network analysis** – These techniques are an apt solution for finding the patterns, groups and perform the link analysis, which is helpful for detection of money laundering, launderer gangs and identify the relationships to find more leads in the money laundering networks. Though there

are challenges in graph mining from a large data processing point of view, but this is one of the potential area where research should be continued.

8. **Ensemble learning** – Considering the number of financial systems in Banks, type of available data, data sources, volume of data; one specific model cannot be a solution to detect suspicious money laundering transactions, instead ensemble learning must be considered to leverage the available data to bring out more confidence in suspiciousness of transaction along with evidences. It could include statistical methods, data mining, machine learning and deep learning all together.

REFERENCES

"16th Web Information Systems and Applications Conference, WISA 2019." In. 2019. *16th Web Information Systems and Applications Conference, WISA 2019*, edited by W. Ni, X. Wang, W. Song and Y. Li. Springer.

Adadi, A., and M. Berrada. 2018. 'Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)', *IEEE Access*, 6: 52138-60.

Al-Shabandar, R., G. Lightbody, F. Browne, J. Liu, H. Wang, and H. Zheng. 2019. "The application of artificial intelligence in financial compliance management." In *2019 International Conference on Artificial Intelligence and Advanced Manufacturing, AIAM 2019*. Association for Computing Machinery.

Alarab, I., S. Prakoonwit, and M. I. Nacer. 2020. "Competence of graph convolutional networks for anti-money laundering in bitcoin blockchain." In, 23-27. Association for Computing Machinery.

Alexandre, C., and J. Balsa. 2016. "Integrating client profiling in an anti-money laundering multi-agent based system." In *World Conference on Information Systems and Technologies, WorldCIST 2016*, edited by M. M. Teixeira, A. M. Correia, H. Adeli, A. Rocha and L. P. Reis, 931-41. Springer Verlag.

Alkhalili, M., M. H. Qutqut, and F. Almasalha. 2021. 'Investigation of Applying Machine Learning for Watch-List Filtering in Anti-Money Laundering', *IEEE Access*, 9: 18481-96.

Alsuwailem, A. A. S., and A. K. J. Saudagar. 2020. 'Anti-money laundering systems: a systematic literature review', *Journal of Money Laundering Control*, 23: 833-48.

AUSTRAC. 2022. 'Typologies and case studies', AUSTRAC, Accessed 20 Mar 2022. https://www.austrac.gov.au/about-us/case-studies.

Badal-Valero, E., J. A. Alvarez-Jareño, and J. M. Pavía. 2018a. 'Combining Benford's Law and machine learning to detect money laundering. An actual Spanish court case', *Forensic Science International*, 282: 24-34.

Badal-Valero, Elena, José A Alvarez-Jareño, and Jose M Pavía. 2018b. 'Combining Benford's Law and machine learning to detect money laundering. An actual Spanish court case', *Forensic Science International*, 282: 24-34.

Baek, H., J. Oh, C. Y. Kim, and K. Lee. 2019. "A Model for Detecting Cryptocurrency Transactions with Discernible Purpose." In *11th International Conference on Ubiquitous and Future Networks, ICUFN 2019*, 713-17. IEEE Computer Society.

Baesens, B., S. Höppner, and T. Verdonck. 2021a. 'Data engineering for fraud detection', *Decision Support Systems*, 150.

Baesens, Bart, Sebastiaan Höppner, and Tim Verdonck. 2021b. 'Data engineering for fraud detection', *Decision Support Systems*, 150: 113492.

Barredo Arrieta, A., N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera. 2020. 'Explainable Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI', *Information Fusion*, 58: 82-115.

Board, Financial Stability. 2017. 'Artificial intelligence and machine learning in financial services '.

Camino, R. D., R. State, L. Montero, and P. Valtchev. 2017. "Finding suspicious activities in financial transactions and distributed ledgers." In *17th IEEE International Conference on Data Mining Workshops, ICDMW 2017*, edited by R. Gottumukkala, G. Karypis, V. Raghavan, X. Wu, L. Miele, S. Aluru, X. Ning and G. Dong, 787-96. IEEE Computer Society.

Canhoto, A. I. 2020. 'Leveraging machine learning in the global fight against money laundering and terrorism financing: An affordances perspective', *Journal of Business Research*.

Cao, D. K., and P. Do. 2012. "Applying data mining in money laundering detection for the Vietnamese banking industry." In *4th Asian Conference on Intelligent Information and Database Systems, ACIIDS 2012*, 207-16. Kaohsiung.

Centre, Australian Transaction Reports and Analysis. 2018. "AUSTRAC and CBA agree $700m penalty." In. https://www.austrac.gov.au/austrac-and-cba-agree-700m-penalty: AUSTRAC.

Centre, Australian Transaction Reports and Analysis. 2020. "AUSTRAC and Westpac agree to proposed $1.3bn penalty." In. https://www.austrac.gov.au/news-and-media/media-release/austrac-and-westpac-agree-penalty: Australian Transaction Reports and Analysis Centre Australia.

Chao, X., G. Kou, Y. Peng, and F. E. Alsaadi. 2019. 'Behavior monitoring methods for trade-based money laundering integrating macro and micro prudential regulation: A case from China', *Technological and Economic Development of Economy*, 25: 1081-96.

Chen, J., T. Ma, and C. Xiao. 2018a. "FastGCN: Fast learning with graph convolu-tional networks via importance sampling." In.: International Conference on Learning Representations, ICLR.

Chen, Tianqi, and Carlos Guestrin. 2016. "Xgboost: A scalable tree boosting system." In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785-94.

Chen, Y., Z. Lu, R. Li, Y. Li, and X. Sun. 2006. 'The research of an incremental conceptive clustering algorithm and its application in detecting money laundering', *Wuhan University Journal of Natural Sciences*, 11: 1076-80.

Chen, Z., W. M. Soliman, A. Nazir, and M. Shorfuzzaman. 2021. 'Variational Autoencoders and Wasserstein Generative Adversarial Networks for Improving the Anti-Money Laundering Process', *IEEE Access*, 9: 83762-85.

Chen, Z., L. D. Van Khoa, E. N. Teoh, A. Nazir, E. K. Karuppiah, and K. S. Lam. 2018b. 'Machine learning techniques for anti-money laundering (AML) solutions in suspicious transaction detection: a review', *Knowledge and Information Systems*, 57: 245-85.

Chen, Z., Khoa Van, L. D.,, E. N. Teoh, A. Nazir, E. K. Karuppiah, and K. S. Lam. 2018c. 'Machine learning techniques for anti-money laundering (AML) solutions in suspicious transaction detection: a review', *Knowledge and Information Systems*, 57: 245-85.

Chollet, François. 2015. "Keras Python library of APIs for deep learning." In. www.keras.io: www.keras.io.

Cirqueira, D., D. Nedbal, M. Helfert, and M. Bezbradica. 2020. "Scenario-Based Requirements Elicitation for User-Centric Explainable AI: A Case in Fraud Detection." In *4th IFIP TC 5, TC 12, WG 8.4, WG 8.9, WG 12.9 International Cross-Domain Conference for Machine Learning and Knowledge Extraction, CD-MAKE 2020*, edited by A. Holzinger, A. Holzinger, P. Kieseberg, A. M. Tjoa, E. Weippl and E. Weippl, 321-41. Springer.

Company, McKinsey &. 2019. 'Risk Transforming approaches to AML and financial crime', McKinsey & Company, Accessed 26th February 2021. https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/Risk/Our%20Insights/Transforming%20approaches%20to%20AML%20and%20financial%20crime/Transforming-approaches-to-AML-and-financial%20crime-vF.pdf.

Crime, United Nations Office on Drugs and. 2021. 'Global Money Laundering', United Nations Office on Drugs and Crime, Accessed 26th February 2021.

Davenport, Tom. 2020. "The Future Of Work Now: AI-Driven Transaction Surveillance At DBS Bank." In *Forbes*. https://www.forbes.com/sites/tomdavenport/2020/10/23/the-future-of-work-now-ai-driven-transaction-surveillance-at-dbs-bank/?sh=4772a6383f7f: Forbes.

Department of Justice, Office of Public Affairs. 2009. 'Credit Suisse Group of Switzerland was fined $536 million', Department of Justice, Office of Public Affairs, Accessed 26th February 2021. https://www.justice.gov/opa/pr/credit-suisse-agrees-forfeit-536-million-connection-violations-international-emergency.

"Digital Forensics and Cyber Crime - First International ICST Conference, ICDF2C 2009, Revised Selected Papers." In. 2010. *1st International Conference on Digital Forensics and Cyber Crime, ICDF2C 2009*. Albany, NY.

Dosilovic, F. K., M. Brcic, and N. Hlupic. 2018. "Explainable artificial intelligence: A survey." In, edited by B. Vrdoljak, E. Tijan, T. G. Grbac, V. Sruk, M. Cicin-Sain, S. Ribaric, K. Skala, M. Koricic, M. Mauher, S. Gros, P. Pale and M. Janjic, 210-15. Institute of Electrical and Electronics Engineers Inc.

Drezewski, R., G. Dziuban, L. Hernik, and M. Paczek. 2016. "Comparison of data mining techniques for Money Laundering Detection System." In *International Conference on Science in Information Technology, ICSITech 2015*, edited by Y. Hendriana, A. Pranolo, A. Prahara and D. P. Ismi, 5-10. Institute of Electrical and Electronics Engineers Inc.

Fan, Jianqing, Cong Ma, and Yiqiao Zhong. 2019. 'A Selective Overview of Deep Learning'.

Farrugia, S., J. Ellul, and G. Azzopardi. 2020. 'Detection of illicit accounts over the Ethereum blockchain', *Expert Systems with Applications*, 150.

Feng, Yixuan, Chao Li, Yun Wang, Jian Wang, Guigang Zhang, Chunxiao Xing, Zhenxing Li, and Zengshen Lian. 2019. "Anti-money laundering (AML) research: a system for identification and multi-classification." In *International Conference on Web Information Systems and Applications*, 169-75. Springer.

Finance, Institute of International. 2018. "Machine Learning in Anti-Money Laundering – Summary Report." In. www.iif.com: Institute of International Finance.

FinCEN. 2022. 'Money laundering case studies', FinCEN, Accessed 20 Mar 2022. https://www.fincen.gov/resources/law-enforcement/case-examples.

Flach, Peter A. 2010. 'ROC Analysis.' in Claude Sammut and Geoffrey I. Webb (eds.), *Encyclopedia of Machine Learning* (Springer US: Boston, MA).

Force, Financial Action Task. 2020. "International Standards On Combating Money Laundering And The Financing Of Terrorism & Proliferation - FATF Recommendations." In, edited by Financial Action Task Force. http://www.fatf-gafi.org/publications/fatfrecommendations/documents/the40recommendationspublishedoctober2004.html: Financial Action Task Force.

Force, Financial Action Task. 2021. 'What is money laundering', Financial Action Task Force, Accessed 26th February 2021. https://www.fatf-gafi.org/faq/moneylaundering/#d.en.11223.

Foundation, Python Software. 1991. 'Python programming language', Python Software Foundation, Accessed 16 Apr 2022. https://www.python.org/.

Fritz-Morgenthal, S., B. Hein, and J. Papenbrock. 2022. 'Financial Risk Management and Explainable, Trustworthy, Responsible AI', *Frontiers in Artificial Intelligence*, 5.

Garcia-Bedoya, O., O. Granados, and J. Cardozo Burgos. 2020. 'AI against money laundering networks: the Colombian case', *Journal of Money Laundering Control*.

George, I., and M. Kavakli. 2010. 'Data mining in the investigation of money laundering and terrorist financing.' in, *Surveillance Technologies and Early Warning Systems: Data Mining Applications for Risk Detection* (IGI Global).

Guevara, J., O. Garcia-Bedoya, and O. Granados. 2020a. "Machine Learning Methodologies Against Money Laundering in Non-Banking Correspondents." In *3rd International Conference on Applied Informatics, ICAI 2020*, edited by H. Florez and S. Misra, 72-88. Springer Science and Business Media Deutschland GmbH.

Guevara, Jorge, Olmer Garcia-Bedoya, and Oscar Granados. 2020b. "Machine learning methodologies against money laundering in non-banking correspondents." In *International Conference on Applied Informatics*, 72-88. Springer.

Gunning, D., and D. W. Aha. 2019. 'DARPA's explainable artificial intelligence program', *AI Magazine*, 40: 44-58.

Gupta, A., D. N. Dwivedi, J. Shah, and A. Jain. 2022. 'Data quality issues leading to sub optimal machine learning for money laundering models', *Journal of Money Laundering Control*, 25: 551-55.

Hall, Patrick, Navdeep Gill, and Nicholas Schmidt. 2019. 'Proposed Guidelines for the Responsible Use of Explainable Machine Learning'.

Han, J., U. Barman, J. Hayes, J. Du, E. Burgin, and D. Wan. 2015a. "NextGen AML: Distributed deep learning based language technologies to augment anti money laundering investigation." In *56th Annual Meeting of the Association for Computational Linguistics, ACL 2018*, 37-42. Association for Computational Linguistics (ACL).

Han, J., U. Barman, J. Hayes, J. Du, E. Burgin, and D. Wan. 2015b. "NextGen AML: Distributed deep learning based language technologies to augment anti money laundering investigation." In, 37-42. Association for Computational Linguistics (ACL).

Han, J., and C. Moraga. 1995. "The influence of the sigmoid function parameters on the speed of backpropagation learning." In *3rd International Workshop on Artificial Neural Networks, IWANN 1995*, edited by J. Mira and F. Sandoval, 195-201. Springer Verlag.

Hepenstal, Sam, Leishi Zhang, Neesha Kodagoda, and B. l william Wong. 2021. 'Developing Conversational Agents for Use in Criminal Investigations', *ACM transactions on interactive intelligent systems*, 11: 1-35.

Huang, G., Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. 2017. "Densely connected convolutional networks." In *30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2261-69. Institute of Electrical and Electronics Engineers Inc.

Huynh, Trung Dong, Niko Tsakalakis, Ayah Helal, Sophie Stalla-Bourdillon, and Luc Moreau. 2021. 'Addressing Regulatory Requirements on Explanations for Automated Decisions with Provenance—A Case Study', *Digital government (New York, N.Y. Online)*, 2: 1-14.

"ICMLT 2020 - Proceedings of 2020 5th International Conference on Machine Learning Technologies." In. 2020. *5th International Conference on Machine Learning Technologies, ICMLT 2020*. Association for Computing Machinery.

Ioffe, S., and C. Szegedy. 2015. "Batch normalization: Accelerating deep network training by reducing internal covariate shift." In *32nd International Conference on Machine Learning, ICML 2015*, edited by F. Bach and D. Blei, 448-56. International Machine Learning Society (IMLS).

Irofti, P., A. Pătraşcu, and A. Băltoiu. 2021. "Fraud detection in networks." In *Studies in Computational Intelligence*, 517-36. Springer Science and Business Media Deutschland GmbH.

Jaiswal, S., and M. Valstar. 2016. "Deep learning the dynamic appearance and shape of facial action units." In.: Institute of Electrical and Electronics Engineers Inc.

James Quinn, Katherine Griffiths. 2009. 'Lloyds Banking Group of UK was fined $350 million (2009)', *The Telegraph,* 10th January 2009.

Jayasree, V., and R. V. Siva Balan. 2015. 'Money laundering identification on banking data using probabilistic relational audit sequential pattern', *Asian Journal of Applied Sciences*, 8: 173-84.

Jayasree, V., and R. V. Siva Balan. 2016. 'Anti money laundering in financial institutions using affiliation mapping calculation and sequential mining', *Journal of Engineering and Applied Sciences*, 11: 51-56.

Journal, The Wall Street. 2012. 'ING Fined a Record Amount, Penalty of $619 Million Tied to Cuba, Iran for Violating U.S. Economic Sanctions', *The Wall Street Journal,* 12 June 2012.

Jullum, M., A. Løland, R. B. Huseby, G. Ånonsen, and J. Lorentzen. 2020. 'Detecting money laundering transactions with machine learning', *Journal of Money Laundering Control*, 23: 173-86.

Kanezashi, H., T. Suzumura, D. Garcia-Gasulla, M. H. Oh, and S. Matsuoka. 2019. "Adaptive Pattern Matching with Reinforcement Learning for Dynamic Graphs." In *25th IEEE International Conference on High Performance Computing, HiPC 2018*, 92-101. Institute of Electrical and Electronics Engineers Inc.

Kim, J., X. Wang, C. Kang, J. Yu, and P. Li. 2021. 'Forecasting air pollutant concentration using a novel spatiotemporal deep learning model based on clustering, feature selection and empirical wavelet transform', *Science of the Total Environment*, 801.

Kingma, Diederik P., and Jimmy Ba. 2014. 'Adam: A Method for Stochastic Optimization'.

Kipf, T. N., and M. Welling. 2017. "Semi-supervised classification with graph convolutional networks." In.: International Conference on Learning Representations, ICLR.

Kitchenham, B., and S. Charters. 2007. 'Guidelines for performing systematic literature reviews in software engineering', *Guidelines for Performing Systematic Literature Reviews in Software Engineering*.

Königstorfer, F., and S. Thalmann. 2020. 'Applications of Artificial Intelligence in commercial banks – A research agenda for behavioral finance', *Journal of Behavioral and Experimental Finance*, 27.

Korauš, A., M. Gombár, P. Kelemen, and S. Backa. 2019. 'Using quantitative methods to identify security and unusual business operations', *Entrepreneurship and Sustainability Issues*, 6: 1101-12.

Kuiper, O., M. van den Berg, J. van der Burgt, and S. Leijnen. 2022. "Exploring Explainable AI in the Financial Sector: Perspectives of Banks and Supervisory Authorities." In *33rd Benelux Conference on Artificial Intelligence, BNAIC/ BENELEARN 2021*, edited by L. A. Leiva, C. Pruski, R. Markovich, A. Najjar and C. Schommer, 105-19. Springer Science and Business Media Deutschland GmbH.

Kuiper, Ouren, Martin van den Berg, Joost van der Burgt, and Stefan Leijnen. 2021. 'Exploring Explainable AI in the Financial Sector: Perspectives of Banks and Supervisory Authorities'.

Kumar, A., S. Das, and V. Tyagi. 2020. "Anti money laundering detection using naïve bayes classifier." In *2020 IEEE International Conference on Computing, Power and Communication Technologies, GUCON 2020*, 568-72. Institute of Electrical and Electronics Engineers Inc.

Kurshan, E., H. Shen, and H. Yu. 2020. "Financial Crime Fraud Detection Using Graph Computing: Application Considerations Outlook." In *2nd International Conference on Transdisciplinary AI, TransAI 2020*, 125-30. Institute of Electrical and Electronics Engineers Inc.

Labanca, D., L. Primerano, M. Markland-Montgomery, M. Polino, M. Carminati, and S. Zanero. 2022. 'Amaretto: An Active Learning Framework for Money Laundering Detection', *IEEE Access*, 10: 41720-39.

Labib, N. M., M. A. Rizka, and A. E. M. Shokry. 2020. "Survey of machine learning approaches of anti-money laundering techniques to counter terrorism finance." In *Lecture Notes in Networks and Systems*, 73-87. Springer.

Le Khac, N. A., and M. T. Kechadi. 2010. "Application of data mining for anti-money laundering detection: A case study." In *10th IEEE International Conference on Data Mining Workshops, ICDMW 2010*, 577-84. Sydney, NSW.

Lecun, Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. 'Deep learning', *Nature (London)*, 521: 436-44.

Leite, G. S., A. B. Albuquerque, and P. R. Pinheiro. 2019. 'Application of technological solutions in the fight against money laundering-A systematic literature review', *Applied Sciences (Switzerland)*, 9.

LexisNexis. 2022. 'Watchlist, PEPs, Sanctions screening', LexisNexis, Accessed 15 Apr 2022. https://risk.lexisnexis.com/global/en/financial-services/financial-crime-compliance/watchlist-screening.

Li, D., L. Sun, X. Xu, Z. Wang, J. Zhang, and W. Du. 2021. 'BLSTM and CNN Stacking Architecture for Speech Emotion Recognition', *Neural Processing Letters*, 53: 4097-115.

Li, X., X. Cao, X. Qiu, J. Zhao, and J. Zheng. 2017. "Intelligent Anti-Money Laundering Solution Based upon Novel Community Detection in Massive Transaction Networks on Spark." In *5th International Conference on Advanced Cloud and Big Data, CBD 2017*, 176-81. Institute of Electrical and Electronics Engineers Inc.

Li, Xiao-Hui ; Cao, Caleb Chen ; Shi, Yuhan ; Bai, Wei ; Gao, Han ; Qiu, Luyu ; Wang, Cong ; Gao, Yuanyuan ; Zhang, Shenjia ; Xue, Xun ; Chen, Lei. 2020. 'A Survey of Data-driven and Knowledge-aware eXplainable AI', *IEEE Transactions on Knowledge and Data Engineering*.

Liang, Y., S. Li, C. Yan, M. Li, and C. Jiang. 2021. 'Explaining the black-box model: A survey of local interpretation methods for deep neural networks', *Neurocomputing*, 419: 168-82.

Liang, Y., M. C. Machado, E. Talvitie, and M. Bowling. 2016. "State of the art control of atari games using shallow reinforcement learning." In, 485-93. International Foundation for Autonomous Agents and Multiagent Systems (IFAAMAS).

Linardatos, P., V. Papastefanopoulos, and S. Kotsiantis. 2021. 'Explainable ai: A review of machine learning interpretability methods', *Entropy*, 23: 1-45.

Liu, R., X. L. Qian, S. Mao, and S. Z. Zhu. 2011. "Research on anti-money laundering based on core decision tree algorithm." In *2011 Chinese Control and Decision Conference, CCDC 2011*, 4322-25. Mianyang.

Liu, X., and P. Zhang. 2010. "A scan statistics based suspicious transactions detection model for Anti-Money Laundering (AML) in financial institutions." In *2010 International Conference on Multimedia Communications, Mediacom 2010*, 210-13. Hong Kong.

Lokanan, M. E. 2019. 'Data mining for statistical analysis of money laundering transactions', *Journal of Money Laundering Control*, 22: 753-63.

Lundberg, S. M., and S. I. Lee. 2017. "A unified approach to interpreting model predictions." In, edited by I. Guyon, R. Fergus, H. Wallach, H. Wallach, I. Guyon, S. V. N. Vishwanathan, U. von Luxburg, R. Garnett, S. V. N. Vishwanathan, S. Bengio and R. Fergus, 4766-75. Neural information processing systems foundation.

Lundberg, Scott. 2018. "SHAP API Library." In. https://shap.readthedocs.io.

Magomedov, S., S. Pavelyev, I. Ivanova, A. Dobrotvorsky, M. Khrestina, and T. Yusubaliev. 2018. 'Anomaly detection with machine learning and graph databases in fraud management', *International Journal of Advanced Computer Science and Applications*, 9: 33-38.

Mark Weber, Jie Chen, Toyotaro Suzumura, Aldo Pareja, Tengfei Ma, Hiroki Kanezashi, Tim Kaler, Charles E. Leiserson, Tao B. Schardl. 2018. 'Scalable Graph Learning for Anti-Money Laundering: A First Look'.

Mashrur, A., W. Luo, N. A. Zaidi, and A. Robles-Kelly. 2020. 'Machine learning for financial risk management: A survey', *IEEE Access*, 8: 203203-23.

Michalak, K., and J. Korczak. 2011. "Graph mining approach to suspicious transaction detection." In *2011 Federated Conference on Computer Science and Information Systems, FedCSIS 2011*, 69-75. Szczecin.

Milind Tiwari, Adrian Gepp and Kuldeep Kumar. 2020. 'A review of money laundering literature: the state of research in key areas', *Pacific Accounting Review*, Vol. 32 pp. 271-303.

Miyake, Sei, and Kunihiko Fukushima. 1984. 'A neural network model for the mechanism of feature-extraction - A self-organizing network with feedback inhibition', *Biological cybernetics*, 50: 377-84.

Molnar, Christoph. 2019. *Interpretable machine learning. A Guide for Making Black Box Models Explainable* (https://christophm.github.io/interpretable-ml-book/).

Monroe, Brian. 2020. 'Fincrime Briefing: AML fines in 2019 breach $8 billion, Treasury official pleads guilty to leaking, 2020 crypto compliance outlook, and more', Association of Certified Financial Crime Specialists, Accessed 26th February 2021. https://www.acfcs.org/fincrime-briefing-aml-fines-in-2019-breach-8-billion-treasury-official-pleads-guilty-to-leaking-2020-crypto-compliance-outlook-and-more/#:~:text=Key%20observations%3A,25%20penalties%20totaling%20%242.29bn.

Nair, V., and G. E. Hinton. 2010. "Rectified linear units improve Restricted Boltzmann machines." In *27th International Conference on Machine Learning, ICML 2010*, 807-14. Haifa.

Network, Financial Crimes Enforcement. 2021. 'Bank Secrecy Act (BSA) Statute', United States Government, Accessed 26th February 2021.

News, CBC. 2012. 'HSBC to pay $1.9B to settle money-laundering case', *CBC News*.

Ngai, E. W. T., Y. Hu, Y. H. Wong, Y. Chen, and X. Sun. 2011. 'The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature', *Decision Support Systems*, 50: 559-69.

Office of Parliamentary Counsel, Canberra, Australia. 2019. "Anti-Money Laundering and Counter-Terrorism Financing Act 2006." In, edited by AUSTRAC. https://www.legislation.gov.au/Details/C2019C00011: AUSTRAC.

Palraj, K., and V. Kalaivani. 2021. 'Predicting the abnormality of brain and compute the cognitive power of human using deep learning techniques using functional magnetic resonance images', *Soft Computing*, 25: 14461-78.

Paula, E. L., M. Ladeira, R. N. Carvalho, and T. Marzagão. 2017a. "Deep learning anomaly detection as suppor fraud investigation in Brazilian exports and anti-money laundering." In, 954-60. Institute of Electrical and Electronics Engineers Inc.

Paula, E. L., M. Ladeira, R. N. Carvalho, and T. Marzagão. 2017b. "Deep learning anomaly detection as suppor fraud investigation in Brazilian exports and anti-money laundering." In *15th IEEE International Conference on Machine Learning and Applications, ICMLA 2016*, 954-60. Institute of Electrical and Electronics Engineers Inc.

Pepe, Margaret Sullivan. 2000. 'Receiver Operating Characteristic Methodology', *Journal of the American Statistical Association*, 95: 308-11.

PricewaterhouseCoopers. 2017. 'Explainable AI Driving business value through greater understanding', PricewaterhouseCoopers (PWC), Accessed 26th February 2021. https://www.pwc.co.uk/services/risk-assurance/insights/explainable-ai.html.

"Proceedings - 2018 IEEE 6th International Conference on Future Internet of Things and Cloud Workshops, W-FiCloud 2018." In. 2018. *6th IEEE International Conference on Future Internet of Things and Cloud Workshops, W-FiCloud 2018*, edited by I. Awan, M. Younas and C. F. Portela. Institute of Electrical and Electronics Engineers Inc.

"Proceedings of the 13th IADIS International Conference ICT, Society and Human Beings 2020, ICT 2020 and Proceedings of the 6th IADIS International Conference Connected Smart Cities 2020, CSC 2020 and Proceedings of the 17th IADIS International Conference Web Based Communities and Social Media 2020, WBC 2020 - Part of the 14th Multi Conference on Computer Science and Information Systems, MCCSIS 2020." In. 2020. *13th IADIS International Conference ICT, Society and Human Beings 2020, ICT 2020 and 6th IADIS International Conference Connected Smart Cities 2020, CSC 2020 and 17th IADIS International Conference Web Based Communities and Social Media 2020,*

WBC 2020, Part of the 14th Multi Conference on Computer Science and Information Systems, MCCSIS 2020. IADIS.

Puiutta, E., and E. M. S. P. Veith. 2020. "Explainable Reinforcement Learning: A Survey." In, edited by A. Holzinger, A. Holzinger, P. Kieseberg, A. M. Tjoa, E. Weippl and E. Weippl, 77-95. Springer.

Rademacher, T. 2019. 'Artificial intelligence and law enforcement.' in, *Regulating Artificial Intelligence* (Springer International Publishing).

Rai, A. 2020. 'Explainable AI: from black box to glass box', *Journal of the Academy of Marketing Science*, 48: 137-41.

Raza, S., and S. Haider. 2011. "Suspicious activity reporting using Dynamic Bayesian Networks." In *1st World Conference on Information Technology, WCIT-2010*, 987-91. Istanbul.

Ribeiro, M. T., S. Singh, and C. Guestrin. 2016. ""Why should i trust you?" Explaining the predictions of any classifier." In, 1135-44. Association for Computing Machinery.

Rocha-Salazar, J. D. J., M. J. Segovia-Vargas, and M. D. M. Camacho-Miñano. 2021. 'Money laundering and terrorism financing detection using neural networks and an abnormality indicator', *Expert Systems with Applications*, 169.

Roshanzamir, A., H. Aghajan, and M. Soleymani Baghshah. 2021. 'Transformer-based deep neural network language models for Alzheimer's disease risk assessment from targeted speech', *BMC Medical Informatics and Decision Making*, 21.

Rudin, C. 2019. 'Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead', *Nature Machine Intelligence*, 1: 206-15.

Sabokrou, M., M. Pourreza, X. Li, M. Fathy, and G. Zhao. 2021. 'Deep-HR: Fast heart rate estimation from face video under realistic conditions', *Expert Systems with Applications*, 186.

Sammut, Claude, and Geoffrey I. Webb. 2010a. 'Accuracy.' in Claude Sammut and Geoffrey I. Webb (eds.), *Encyclopedia of Machine Learning* (Springer US: Boston, MA).

Sammut, Claude, and Geoffrey I. Webb. 2010b. 'Area Under Curve.' in Claude Sammut and Geoffrey I. Webb (eds.), *Encyclopedia of Machine Learning* (Springer US: Boston, MA).

Sammut, Claude, and Geoffrey I. Webb. 2010c. 'F1-Measure.' in Claude Sammut and Geoffrey I. Webb (eds.), *Encyclopedia of Machine Learning* (Springer US: Boston, MA).

Sammut, Claude, and Geoffrey I. Webb. 2010d. 'False Negative.' in Claude Sammut and Geoffrey I. Webb (eds.), *Encyclopedia of Machine Learning* (Springer US: Boston, MA).

Sammut, Claude, and Geoffrey I. Webb. 2010e. 'False Positive.' in Claude Sammut and Geoffrey I. Webb (eds.), *Encyclopedia of Machine Learning* (Springer US: Boston, MA).

Sammut, Claude, and Geoffrey I. Webb. 2010f. 'Recall.' in Claude Sammut and Geoffrey I. Webb (eds.), *Encyclopedia of Machine Learning* (Springer US: Boston, MA).

Sammut, Claude, and Geoffrey I. Webb. 2010g. 'True Negative.' in Claude Sammut and Geoffrey I. Webb (eds.), *Encyclopedia of Machine Learning* (Springer US: Boston, MA).

Sammut, Claude, and Geoffrey I. Webb. 2010h. 'True Positive.' in Claude Sammut and Geoffrey I. Webb (eds.), *Encyclopedia of Machine Learning* (Springer US: Boston, MA).

Savage, D., X. Zhang, Q. Wang, X. Yu, and P. Chou. 2017a. "Detection of money laundering groups: Supervised learning on small networks." In *31st AAAI Conference on Artificial Intelligence, AAAI 2017*, 43-49. AI Access Foundation.

Savage, D., X. Zhang, Q. Wang, X. Yu, and P. Chou. 2017b. "Detection of money laundering groups: Supervised learning on small networks." In, 43-49. AI Access Foundation.

Schmidhuber, J. 2015. 'Deep Learning in neural networks: An overview', *Neural Networks*, 61: 85-117.

Semenov, A., A. Mazeev, D. Doropheev, and T. Yusubaliev. 2017. "Survey of common design approaches in AML software development." In, edited by A. Simonov and V. Voevodin, 1-9. CEUR-WS.

Senator, Ted E, Henry G Goldberg, Jerry Wooton, Matthew A Cottini, AF Umar Khan, Christina D Klinger, Winston M Llamas, Michael P Marrone, and Raphael WH Wong. 1995. 'Financial crimes enforcement network AI system (FAIS) identifying potential money laundering from reports of large cash transactions', *AI Magazine*, 16: 21-21.

Shah, R. S., A. Bhatia, A. Gandhi, and S. Mathur. 2021. "Bitcoin Data Analytics: Scalable techniques for transaction clustering and embedding generation." In *2021 International Conference on COMmunication Systems and NETworkS, COMSNETS 2021*. Institute of Electrical and Electronics Engineers Inc.

Shaju, B., and N. Valliammal. 2019. 'Measures for financial fraud detection using data analytics and machine learning', *International Journal of Advanced Science and Technology*, 28: 270-80.

Shapley, L. S. 1953. 'A Value for n-Person Games', *A Value for N-Person Game*: 307-17.

Shokry, Amr Ehab Muhammed, Mohammed Abo Rizka, and Nevine Makram Labib. 2020. "Counter terrorism finance by detecting money laundering hidden networks using unsupervised machine learning algorithm." In *International Conferences ICT, Society, and Human Beings*.

Shrikumar, A., P. Greenside, and A. Kundaje. 2017. "Learning important features through propagating activation differences." In *34th International Conference on Machine Learning, ICML 2017*, 4844-66. International Machine Learning Society (IMLS).

Silva, J. C. S., D. MacEdo, C. Zanchettin, A. L. I. Oliveira, and A. T. De Almeida Filho. 2021. "Multi-Class Mobile Money Service Financial Fraud Detection by Integrating Supervised Learning with Adversarial Autoencoders." In *2021 International Joint Conference on Neural Networks, IJCNN 2021*. Institute of Electrical and Electronics Engineers Inc.

Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. 2014. 'Dropout: A simple way to prevent neural networks from overfitting', *Journal of Machine Learning Research*, 15: 1929-58.

State of California, USA. 2018. "TITLE 1.81.5. California Consumer Privacy Act of 2018." In.

Stepin, I., J. M. Alonso, A. Catala, and M. Pereira-Farina. 2021. 'A Survey of Contrastive and Counterfactual Explanation Generation Methods for Explainable Artificial Intelligence', *IEEE Access*, 9: 11974-2001.

Sudjianto, A., S. Nair, M. Yuan, A. Zhang, D. Kern, and F. Cela-Díaz. 2010. 'Statistical methods for fighting financial crimes', *Technometrics*, 52: 5-19.

Systems, BAE. 2022. 'NetReveal Sanctions and PEP Screening', BAE Systems, Accessed 15 Apr 2022. https://www.baesystems.com/en-financialservices/solutions/banking-compliance/sanctions-and-pep-screening.

Szegedy, C., W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. 2015. "Going deeper with convolutions." In, 1-9. IEEE Computer Society.

Tang, J., and J. Yin. 2005. "Developing an intelligent data discriminating system of anti-money laundering based on SVM." In, 3453-57.

Team, Google Brain. 2015. "TensorFlow - An end-to-end open source machine learning platform." In. https://www.tensorflow.org/install.

Tertychnyi, P., I. Slobozhan, M. Ollikainen, and M. Dumas. 2020. "Scalable and Imbalance-Resistant Machine Learning Models for Anti-money Laundering: A Two-Layered Approach." In *10th International Workshop on Enterprise Applications, Markets and Services in the Finance Industry, FinanceCom 2020*, edited by B. Clapham and J. Koch, 43-58. Springer Science and Business Media Deutschland GmbH.

Ting, Kai Ming. 2010a. 'Precision.' in Claude Sammut and Geoffrey I. Webb (eds.), *Encyclopedia of Machine Learning* (Springer US: Boston, MA).

Ting, Kai Ming. 2010b. 'Precision and Recall.' in Claude Sammut and Geoffrey I. Webb (eds.), *Encyclopedia of Machine Learning* (Springer US: Boston, MA).

Tjoa, E., and C. Guan. 2021. 'A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI', *IEEE Transactions on Neural Networks and Learning Systems*, 32: 4793-813.

Torres, R. A. L., and M. Ladeira. 2020. "A proposal for online analysis and identification of fraudulent financial transactions." In *19th IEEE International Conference on Machine Learning and Applications, ICMLA 2020*, edited by M. A. Wani, F. Luo, X. Li, D. Dou and F. Bonchi, 240-45. Institute of Electrical and Electronics Engineers Inc.

Townsend, J., T. Chaton, and J. M. Monteiro. 2020. 'Extracting Relational Explanations from Deep Neural Networks: A Survey from a Neural-Symbolic Perspective', *IEEE Transactions on Neural Networks and Learning Systems*, 31: 3456-70.

Tundis, A., S. Nemalikanti, and M. Mühlhäuser. 2021. "Fighting organized crime by automatically detecting money laundering-related financial transactions." In *16th International Conference on Availability, Reliability and Security, ARES 2021*. Association for Computing Machinery.

Union, European. 2016. "General Data Protection Regulation (GDPR)." In.: Official Journal of the European Union.

Wang, Su-Nan, and Jian-Gang Yang. 2007. "A money laundering risk evaluation method based on decision tree." In *2007 International conference on machine learning and cybernetics*, 283-86. IEEE.

Westerski, A., R. Kanagasabai, E. Shaham, A. Narayanan, J. Wong, and M. Singh. 2021. 'Explainable anomaly detection for procurement fraud identification—lessons from practical deployments', *International Transactions in Operational Research*.

Xu, F., H. Uszkoreit, Y. Du, W. Fan, D. Zhao, and J. Zhu. 2019. "Explainable AI: A Brief Survey on History, Research Areas, Approaches and Challenges." In, edited by J. Tang, M. Y. Kan, D. Zhao, S. Li and H. Zan, 563-74. Springer.

Yang, Q., B. Feng, and P. Song. 2007. "Study on anti-money laundering service system of online payment based on union-bank mode." In *2007 International Conference on Wireless Communications, Networking and Mobile Computing, WiCOM 2007*, 4986-89. Shanghai.

Yang, Y., B. Lian, L. Li, C. Chen, and P. Li. 2014. "DBSCAN clustering algorithm applied to identify suspicious financial transactions." In *6th International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery, CyberC 2014*, 60-65. Institute of Electrical and Electronics Engineers Inc.

Yasaka, N. 2020. 'Global knowledge management of suspicious transaction reporting system in Japan', *Journal of Money Laundering Control*, 23: 55-63.

Zand, A., J. Orwell, and E. Pfluegel. 2020. "A Secure Framework for Anti-Money-Laundering using Machine Learning and Secret Sharing." In *2020 International Conference on Cyber Security and Protection of Digital Services, Cyber Security 2020*. Institute of Electrical and Electronics Engineers Inc.

Zhang, C. H., and X. H. Zhao. 2008. 'Research on money laundering recognition based on decision thee algorithm', *Wuhan Ligong Daxue Xuebao/Journal of Wuhan University of Technology*, 30: 154-56.

Zhang, Dongsong, and Lina Zhou. 2004. 'Discovering golden nuggets: data mining in financial application', *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 34: 513-22.

Zhang, Y., and P. Trubey. 2019. 'Machine Learning and Sampling Scheme: An Empirical Study of Money Laundering Detection', *Computational Economics*, 54: 1043-63.

Zhang, Z., J. J. Salerno, and P. S. Yu. 2003. "Applying data mining in investigating money laundering crimes." In *9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '03*, 747-52. Washington, DC.

Zhdanova, M., J. Repp, R. Rieke, C. Gaber, and B. Hemery. 2014. "No smurfs: Revealing fraud chains in mobile money transfers." In *9th International Conference on Availability, Reliability and Security, ARES 2014*, 11-20. Institute of Electrical and Electronics Engineers Inc.

APPENDIX A – MACHINE LEARNING HYPER PARAMETERS

Hyper Parameters – RF

- n_estimators = 100,
- criterion = 'gini',
- min_samples_split = 2,
- min_samples_leaf = 1,
- min_weight_fraction_leaf = 0.0
- max_features = 'sqrt',
- min_impurity_decrease = 0.0,
- bootstrap = True,
- oob_score = False,
- verbose = 0,
- warm_start = False,
- class_weight = (0:1,1:100),
- ccp_alpha = 0.0

Hyper Parameters – SVM

- C = 1.0,
- Kernel = 'linear',
- Degree = 3,
- Gamma = 'scale',
- coef0 = 0.0,
- shrinking = True,
- probability = False,
- tol = 0.001,
- cache_size = 200,
- verbose = False,
- max_iter = -1,
- decision_function_shape = 'ovr',

- break_ties = False.

Hyper Parameters – XGBoost

- lLoss = 'log_loss',
- learning_rate = 0.1,
- n_estimators = 100,
- subsample = 1.0,
- criterion = 'friedman_mse',
- min_samples_split = 2,
- min_samples_leaf = 1,
- min_weight_fraction_leaf = 0.0,
- max_depth = 3,
- min_impurity_decrease = 0.0,
- verbose = 0,
- warm_start = False,
- validation_fraction = 0.1,
- tol = 0.0001,
- ccp_alpha = 0.0.

APPENDIX B – INTERPRETATION EXAMPLES USING SHAP

Transactions predicted as suspicious



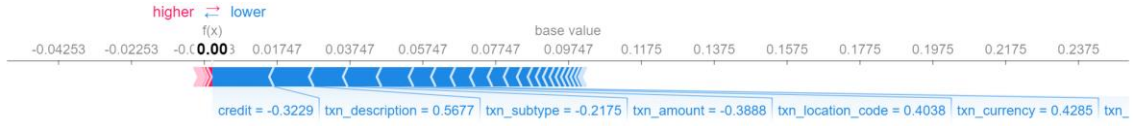


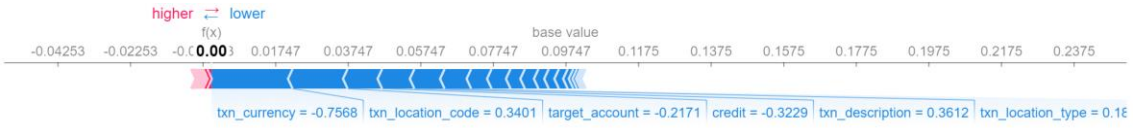


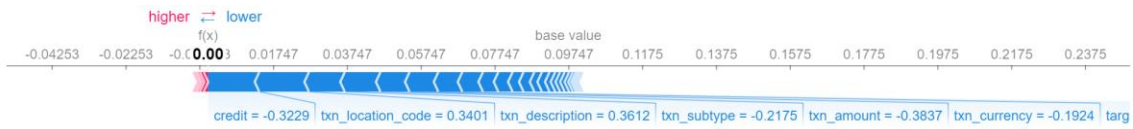




Transactions predicted as legitimate









APPENDIX C – KEY TERMINOLOGIES

Terminology	Definition
Money laundering	Money laundering is a criminal activity performed to disguise the illegally obtained money and make them appear legitimate in the system.
Anti-money laundering	Anti-money laundering is a set of activities performed by reporting entities to monitor the transactions and report the suspicious activities to regulatory.
Terrorism financing	Terrorism financing is a process of financing the terrorist groups or terrorist organisations to perform illegal activities.
Anti-terrorism financing	Anti-terrorism financing is a set of activities performed to prevent terrorism financing.
Artificial intelligence	A development of theory and computer systems that are able to perform the tasks that requires human intelligence.
Machine learning	Machine learning is an application of computer algorithm to learn from the data, predict certain outcomes and improve from experience from without being programmed.
Deep learning	Deep learning is a machine learning technique that trains the machine to predict the outcomes the way human brain thinks.
Data mining	Data mining is process of discovering the patterns and valuable information from the large dataset using several methods including statistical methods, machine learning methods and database systems.
Explainable artificial intelligence	Explainable artificial intelligence is a process and method that helps the humans to comprehend and trust the decisions made by machine learning methods.
Responsible artificial intelligence	Responsible artificial intelligence is an approach of designing and deploying artificial intelligence solutions using machine learnings with good intentions that produces unbiased solutions, empower businesses, safeguards the privacy, instil the trust and confidence in the system.

Terminology	Definition
Explainability	Explainability refers to the fact if humans are able to comprehend the predictions made by machine learning models by looking at the input data.
Interpretability	Interpretability refers to how accurately a machine learning model can associate a cause to an effect.
Model transparency	Model transparency refers to how much humans can understand how machine learning model is functioning internally to predict the results.

APPENDIX D – DATA GENERATION PROCESS

