sciendo INTERNATIONAL JOURNAL ON SMART SENSING AND INTELLIGENT SYSTEMS

Article | DOI: 10.2478/ijssis-2024-0027 Issue 1 | Vol. 17 (2024)

Explainable deep learning model for predicting money laundering transactions

Dattatray Vishnu Kute[1], Biswajeet Pradhan[1,*], Nagesh Shukla[2] and Abdullah Alamri[3]

[1]Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, NSW 2007, Australia

[2]Department of Business Strategy and Innovation, Griffith Business School, Griffith University Brisbane, Queensland 4111, Australia

[3]Department of Geology and Geophysics, College of Science, King Saud University, Riyadh 11451, Saudi Arabia

*E-mail: Biswajeet.pradhan@uts.edu.au/biswajeet24@gmail.com

Received for publication April 10, 2024.

Abstract

Money laundering has been a global issue for decades. The ever-changing technology landscape, digital channels, and regulations make it increasingly difficult. Financial institutions use rule-based systems to detect suspicious money laundering transactions. However, it suffers from large false positives (FPs) that lead to operational efforts or misses on true positives (TPs) that increase the compliance risk. This paper presents a study of convolutional neural network (CNN) to predict money laundering and employs SHapley Additive exPlanations (SHAP) explainable artificial intelligence (AI) method to explain the CNN predictions. The results highlight the role of CNN in detecting suspicious transactions with high accuracy and SHAP's role in bringing out the rationale of deep learning predictions.

Keywords

deep learning, explainable AI, financial crime, money laundering, SHAP, suspicious transaction, CNN

I. Introduction

Money laundering has been a global issue for decades and is an ongoing threat for economies and societies. The United Nations Office on Drugs and Crime (UNODC) estimates that the amount of money laundered globally every year is between 2% and 5% of global gross domestic products (GDP) value which exceeds 1 trillion dollars [1]. Money laundering is a process of disguising the original source of funds obtained from illicit activities and legitimizing them through the financial system. The financial action task force (FATF), an intergovernmental organization formed in 1989, has published a standard set of recommendations that governments should follow to combat money laundering and terrorist financing. As per the FATF recommendations [2], most governments have established the regulations and policies to combat money laundering and terrorism financing. For example, the United States has established The Bank Secrecy Act [3]; Australia has established anti-money laundering and countering terrorism financing (AML/CTF) Act [4]; and India has established Prevention of Money Laundering Act [5]. Most countries have set up the financial intelligence units (FIUs) [6–8] that are responsible for receiving, processing, and analyzing suspicious activity reports submitted by regulated entities to identify money laundering, terrorism financing, tax evasion, and other financial criminal activities. Regulated entities include the organizations involved in cash business, financial services, bullion, and cryptocurrencies. Each regulated entity is obliged to ensure regulatory compliance, and one of the key compliance requirements is to report suspicious transactions to regulatory authorities. Failing to do so attracts heavy penalties, damages brand

reputation, and ultimately leads to loss of business. Most recent examples are as follows: in year 2023, the United States Department of Justice charged Binance and its CEO with US $4.3 billion for non-compliance with AML laws and sanctions regulations [9]; in year 2022, the largest bank in Denmark was involved in one of the world's biggest money laundering scandals and was forfeited with US $2.0 billion [10]; and in year 2020, one of the Big Four Banks in Asia-Pacific region was issued a penalty of AUD $1.3 billion for breach of AML/CTF policy [11]. In 2019, approximately 58 AML penalties were issued, totaling US $8 billion, which is two times larger than the amount issued in 2018 [12].

Based on research commissioned by Refinitiv in 2018 and conducted by an independent third party involving 2,300 senior C-suite executives (from companies with average annual turnover of US $17.4 billion) from 19 countries, the true cost of combating the financial crime was identified as 3.1% of annual turnover which is equivalent to AUD $1.28 trillion [13]. To combat money laundering, financial institutions use transaction monitoring and rule-based AML systems. These systems raise alerts for suspicious transactions, international fund transfers (IFT), and cash transactions beyond threshold amounts. Suspicious transaction alerts are further investigated by AML experts who determine if the transactions identified in the alerts are indeed suspicious (also called as true positives [TPs]) or false alerts (also called as false positives [FPs]). The FP rate of these alerts is estimated over 98% [14]. If there is adequate evidence to qualify the transactions as suspicious, the same is reported by preparing the suspicious matter report (SMR) to regulatory authority. According to the EU's law enforcement agency Europol, out of every 1,000 transactions flagged by Banks to FIUs, only 50 transactions are referred to law and enforcement, and only five of these lead to criminal investigations [13]. One can imagine the manual efforts, number of staff members, and cost required to validate the humongous number of alerts each day, a number that continues to grow with the advancements in digital banking.

Considering the continuous increase in transaction volumes, changing regulatory and technological landscapes, and evolving fraud patterns, few technology-savvy financial institutions have started experimenting with artificial intelligence/machine learning (AI/ML)-based AML systems to identify suspicious transactions. However, they face several challenges when attempting to drive the adoption of AI-based systems. These challenges come from the lack of interpretability, transparency, and explainability of decisions made using opaque models. This creates difficulty in answering questions regarding the auditability of systems, trust in the system, data privacy, ethical usage of the data, and bias in decision making [15, 16]. According to the annual global CEO survey conducted by PWC in 2020, involving 1,378 chief executives from >90 countries, 84% of CEOs believe that AI-based decisions need to be explained in order to be trusted [17].

Regulations such as the general data protection regulation (GDPR) [18] and California consumer private act (CCPA) [19] have mandated the explainability of AI/ML-based solutions in the regulatory compliance domain. The regulator demands sufficient evidence along with the SMR to justify the suspiciousness of transactions. Each SMR submitted by a bank conveys a message to the regulatory that a customer associated with the reported suspicious transactions requires further investigation. The outcome of the SMR review by the regulatory can lead to criminal investigation of the customer. Hence, one of the top priorities of banks is to ensure that genuine customers are not falsely reported while simultaneously preventing fraudsters from exploiting the financial system for money laundering. As per the systematic literature review [20], the research on the application of deep learning methods in the AML domain is limited and there is no evidence of XAI techniques being applied to explain the decisions made by deep learning methods.

a. Key challenges and motivations

Despite the concerted efforts to combat the money laundering by intergovernmental organizations, national government agencies, and law and enforcement, regulatory, financial institutions, the global financial system is still being exploited for money laundering. Financial institutions, which are the first line of defense to identify suspicious transactional behavior, hold authoritative positions to contribute effectively. This study has identified the following challenges for financial institutions in the money laundering domain.

1. High FP alerts raised by rule-based AML systems for suspicious transactions lead to an increase in operational costs and reduce operational efficiencies.
2. High false negatives (FNs) cause true money laundering transactions to go unnoticed, leading to an

increase in compliance risk and continuing criminal activities that negatively impact society.
3. The lack of explainability by highly accurate deep learning techniques hinders the adoption of AI technology in the AML domain.

Hence, the objectives of this study are to address these challenges by designing an effective deep learning technique to detect suspicious money laundering transactions with minimal FPs and FNs and to develop an effective explainable AI technique to produce human interpretable explanations.

b. Proposed methodology

To address these objectives, a novel method to detect suspicious transactions using a Conv1D convolutional neural network (CNN) was implemented and explained the CNN predictions using the state-of-the-art SHapley Additive exPlanations (SHAP) XAI method, which is based on the approach outlined by Kute [21]. The CNN results were compared with three other most commonly used ML methods for detecting money laundering: random forest (RF) [22], extreme gradient boosting (XGBoost) [23], and support vector machine (SVM) [24]. This study chose Conv1D CNN model for detecting suspicious transactions because of its suitability for sequential financial time-series data and tabular dataset. A model called agnostic method SHAP was chosen to explain the predictions made by CNN primarily due to its ability to provide an individual feature importance score for each input feature value, indicating contribution to prediction. All classifiers used in this study were trained, validated, and tested on synthetic data generated by the authors.

c. Key contributions of this study

1. This is the first study to use a Conv1D CNN to detect suspicious money laundering transactions in the banking domain.
2. This study employs the SHAP XAI method to explain suspicious transaction predictions made by CNN, which, to our knowledge, is a novel approach.

d. Significance of this study

A bank or any regulated financial institution can potentially consider the methodology explained in this study to detect suspicious customer transactional behavior. The explainability part of the methodology can help the investigator interpret the decisions made by opaque models. Detecting and preventing money laundering brings several benefits to society which include mitigating the harmful effects of criminal activity, reducing terrorism financing, maintaining the integrity of financial systems, and protecting the economy. Money laundering is often associated with criminal activities such as drug trafficking, human trafficking, organized crimes, and corruption. Therefore, the detection of money laundering can assist law enforcement agencies in identifying criminal organizations, disrupting organized crimes, preventing illicit activities, and ultimately contributing to the safety and security of society.

The remaining paper is organized as follows: Section 2 describes the related work from the literature; Section 3 describes the methodology, including synthetic data generation, deep learning method, and XAI method; Section 4 provides the results; Section 5 discusses the results and provides a viewpoint; and finally, Section 6 concludes the paper by providing key findings and future research directions.

II. Related Work

The detection of suspicious money laundering transactions using statistical and ML methods is a fairly well-researched area over the past couple of decades [25], but with a limited research using deep learning techniques [20]. Previous work in the AML domain using deep learning includes a scalable graph convolutional neural network (GCN) for forensic analysis of financial data to provide visual analysis as part of decision support systems used by AML analysts [26]. The study was conducted using synthetic data. Another study used a GCN together with a multilayer perceptron classifier to predict illicit transactions in a Bitcoin transaction graph using a publicly available elliptic dataset of real Bitcoin transactions [27].

The suspicious transaction alerts generated by rule-based system was used as inputs for deep learning models (Natural Language Processing-driven multichannel convolution neural networks) to perform sentiment and link analyses (using diverse set of data such as news, Twitter, and social media), presenting the results in visual form [28]. Autoencoder classifiers were used to identify possible fraudsters through anomaly detection by analyzing the transaction patterns in Brazil's foreign trade database containing 50,000 legal entities in shipping goods to 200 countries [29]. More recent studies include a dynamic graph attention network to detect suspicious

accounts involved in illicit activities [30], the application of recurrent and transformer encoder models to raise AML alarms [31], the use of HAMLET, a scalable deep learning model for detecting money laundering patterns [32], the adoption of a graph neural network known as node and edge neural network (NENN) to improve the decision-making ability of AML systems [33], the implementation of a group-aware deep graph learning-based approach for organized money laundering detection [34], the use of a meta path encoder for detecting Bitcoin money laundering transactions [35], and the performance of transactional network analysis using graph convolutional network and recurrent neural network to detect money laundering behavior [36].

Owing to the development of *post hoc* methods, deep learning has garnered increased attention in the field of money laundering detection, where explainability of decisions is most important along with the detection accuracy. This explainability of predictions is crucial for AML Investigation Officers to validate the decisions made by the deep learning classifier and gather evidence to build a case for further investigation. An explanation can be either a visual or textual representation of the connection between input features used for prediction [37]. A more comprehensive explanation of a prediction helps the AML Investigation Officer make a quick decision if the identified transaction(s) is legitimate or adequately suspicious so that it can be considered for the next level of investigation. In turn, this can significantly improve operational efficiency. Hence, there is an immediate need for the application of explainable deep learning methods in the finance domain to help remove the AI adoption barrier.

III. Data and Methodology

The methodology is designed by considering the following two goals: (1) predicting suspicious money laundering transactions and (2) explaining the predictions by showing the most influential features that contribute to a decision. Figure 1 shows the design of the research development methodology that was used to conduct the research presented in this article. This methodology has four major parts: Part 1 focuses on synthetic data generation containing customers, accounts, and transactions in tabular format; Part 2 focuses on developing a CNN classifier based on Conv1D layers for predicting suspicious transactions; Part 3 focuses on explaining the predictions made by CNN using SHAP XAI techniques; and Part 4 focuses on comparing the prediction performance of the CNN classifier with the RF, XGBoost, and SVM ML classifiers.

Typically, banks consider customer data, accounts or products, transaction list, watch lists, and sanction list as key data entities to detect suspicious transactions. It is relatively easy to detect money laundering transactions more effectively at the *placement* stage when money is deposited in a savings account. It becomes more difficult from the *structuring* stage, which includes transfer type of transactions to other savings accounts, credit cards, loan accounts, insurance accounts, etc., and is extremely difficult at the *integration* stage, where money is integrated into the economy by investing in legitimate businesses. Considering the complexities of transaction creation for each type of product (e.g., savings account, credit card, loan account, and insurance) and then integrating them together, we limited the scope of synthetic data creation to key data entities—customer, savings account, and transactions—that were represented using 40 attributes. Customer creation rules were defined to ensure a balanced spread across geography and other demographic parameters. Transaction creation rules were defined to ensure basic finance rules and to keep transactions as real as possible. Money laundering scenarios were identified to generate transactions and label them as suspicious to train the ML model. A software was developed to automatically generate data and save it in the database. The generated data were exported to Excel and manually ingested with suspicious transactions. The data underwent several rounds of generation and refinement, including adding noise to the data.

After creating synthetic data, a model using a CNN classifier was developed, trained, and tested on the generated dataset. The model has undergone several rounds of tuning the hyperparameters and testing to achieve the best performance and finalize the architecture of the CNN model. The RF, XGBoost, and SVM machine learning models were chosen to compare the performance of the CNN model. The ML models were trained and tested by adjusting the hyperparameters on the same dataset that was used to train and test the CNN model.

Following the development of the CNN model, the SHAP method was used to explain the CNN predictions. The authors developed a software code to apply SHAP to the CNN classifier, generate Shapley values, and explain the predictions made by the CNN using multiple graphs. Local explanations were provided at the individual record prediction level, highlighting the features that positively or negatively

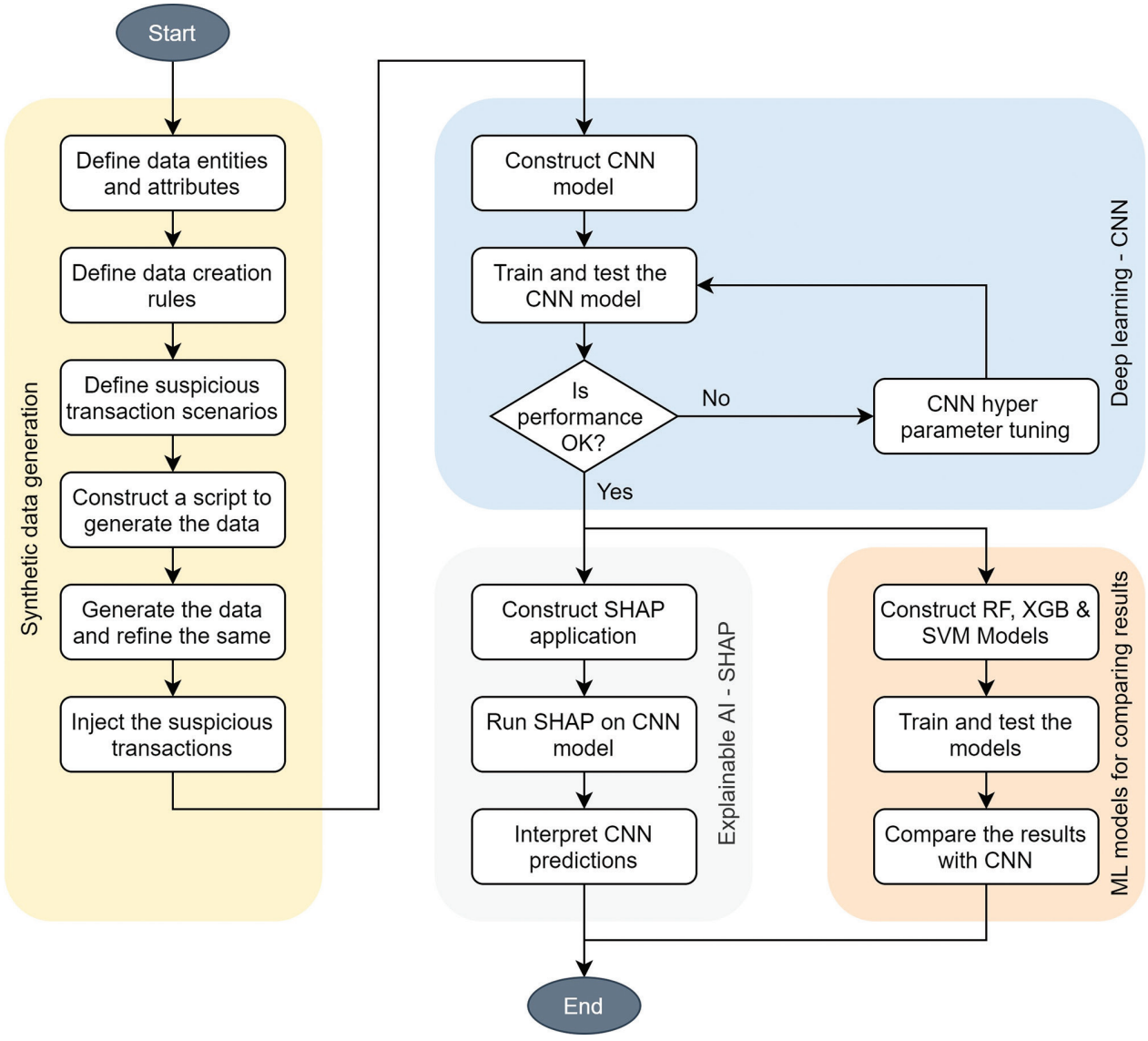


Figure 1: Research development methodology for generating synthetic data, predicting money laundering transactions using CNN, and interpreting the predictions using SHAP. AI, artificial intelligence; CNN, convolutional neural network; ML, machine learning; RF, random forest; SHAP, SHapley Additive exPlanations; SVM, support vector machine.

influenced the decisions. Further details are provided in the following sections.

a. Synthetic data generation

This research requires financial transaction data together with labeled money laundering transactions. Banks consider financial transaction data, including customer, accounts, history, balance, and products, as highly sensitive and protected data, which makes it difficult to obtain the same for research purposes. It is also important to note that, even in real banking transaction data, no one knows for sure if the transactions are legitimate or money laundering transactions until it is proved in the court; hence, the labeled data for training the ML models is also a challenge for banks. The decision points to determine if the identified transactions are indeed suspicious or legitimate are spread across several entities such as banks, FIU, law and enforcement, and court.

To our knowledge, real financial transaction dataset of a bank does not exist on public domain that provides labeled transaction data for research

purposes (and it should not be as a responsible bank). Considering the challenges surrounding the availability of financial transaction data for research, it was decided to produce synthetic data that would be as close as possible to the regular transactions of retail customers of a bank. The following key data entities were considered: customers, accounts, and transactions for data creation. Figure 2 illustrates the synthetic data creation process.

The attributes for the data entities were chosen based on the real banking customer transactions dataset and the attributes used by other researchers [25] in the same domain. Each data attribute was carefully chosen, considering its potential contribution to determine whether the transaction is suspicious or legitimate. Additionally, these attributes were validated by AML subject matter experts (SMEs) to ensure that the correct data were fed into the model for improved decision making.

The key attributes considered for *customer* include customer ID, customer type, gender, date of birth, age, marital status, residence country, state, city, postcode, tax resident country, birth country, nationality country, profession, income category, know your customer (KYC) update date, KYC state, risk rating, and account number.

The attributes for *account* entity are chosen as account number, customer ID, bank state branch (BSB) number, account creation date, account type, daily transaction limit, tax file number (TFN) number, and account statement delivery method.

The attributes for *transaction* entity are chosen as transaction date, transaction number, source account number, amount, credit, debit, transaction type,

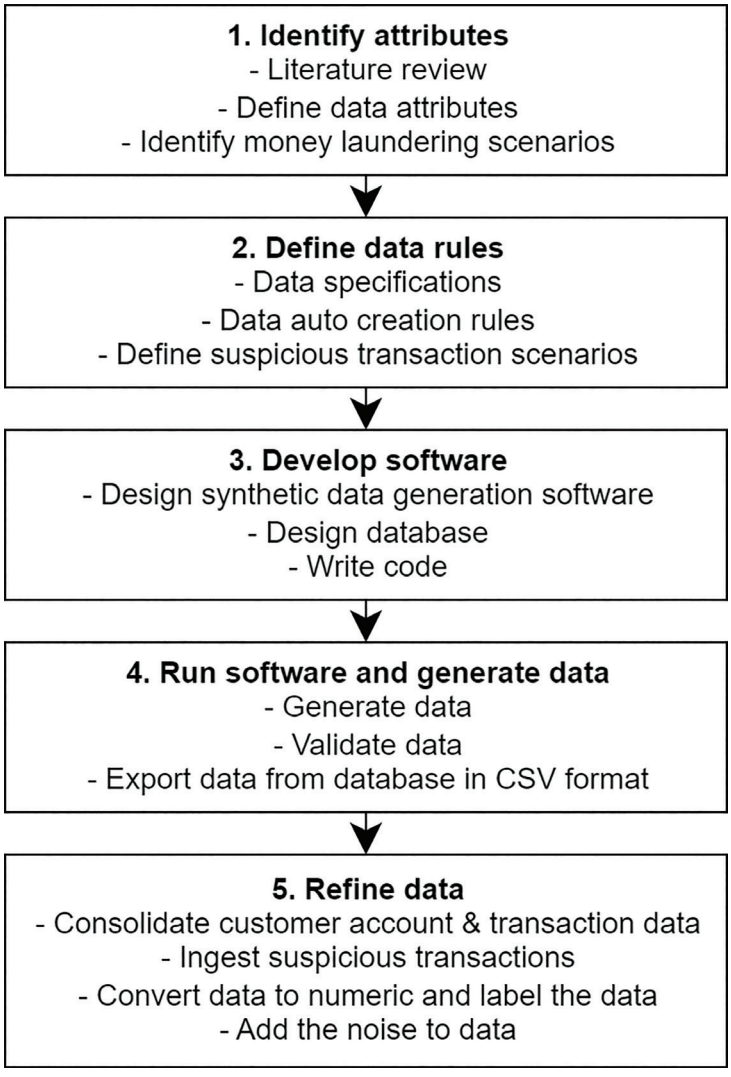


Figure 2: Synthetic financial transaction data generation methodology.

subtype, description, currency, transaction location type, transaction location code, target account number, target country code, target bank code, and is suspicious flag.

Demographic information from the Australian Bureau of Statistics [38] was used to maintain a realistic balance between customer profiles, age, city, profession, gender, and income. Customer profiles were created by limiting a country of residence to Australia, 12 cities, age group between 25 years and 60 years containing males and females; standard occupations such as managers, technicians, and trade workers, community and personal service workers, clerical and administrative workers, sales workers, machinery operators and drivers, laborers, and students; and an annual income range between AU $40,000 and $150,000.

Each customer profile was associated with one savings account. Transactions on the savings account were generated by meticulously defining the ranges of transaction dates and transaction amounts based on customer profiles, and the purpose of the transaction to keep the data as close as possible to real transactions. Credit transactions on the account include salary, cash deposits, and incoming transfers. Debit transactions include rent, home loan EMI, energy bills, phone bills, vehicle loan EMI, health insurance, vehicle registration, green slip, vehicle insurance, school fees, cash withdrawal, payment at counters, money transfer, bills, and shopping. To maintain a realistic number of transactions per customer, we used a combination of a fixed set of transactions and a few random transactions such as shopping and restaurant expenses. Transaction consistency and integrity is maintained by ensuring that the account balance does not go negative. Financial transaction data are inherently unbalanced, with most transactions being legitimate and a minor number of transactions being suspicious. In practice, this ratio is typically approximately 99:1 with slight variations. To balance the synthetic data, we chose to label 95.64% of transactions as legitimate and 4.36% as suspicious. Money laundering transactions were created based on the suspicious money laundering scenarios listed in Table 1. In the banking sector, suspicious transaction data originate from alerts generated by rule-based AML systems, cases prepared post-alert cases, and SMRs submitted to regulatory authorities.

Table 2 shows some legitimate transaction scenarios that look like suspicious money laundering transactions, which are also called *overlapping* transactions. Such transactions are also used to train and verify whether the model can identify true money laundering transactions.

b. CNN model for money laundering detection

CNN is one of the methods of deep learning [39] that is widely and successfully used in the computer vision domain [40] for image analysis [41], classification and identification, speech recognition [42], and natural language processing [43]. CNN is also one of the most studied algorithms to address

Table 1: Scenarios to develop money laundering transactions

S. No.	Scenario description
ML-1	Small deposits (<AU $5,000) of money through ATM by multiple people into a single account (<AU $10,000 per day) over a month. Then the same money is transferred in batches of AU $10,000 to $30,000 to multiple overseas accounts in different countries.
ML-2	Small deposits (<AU $5,000) of money through ATM by multiple people into a single account (<AU $10,000 per day) over a month. Then the same money is used to buy luxurious items locally in the range of AU $10,000 to AU $90,000 (vehicles, gold, property, etc.).
ML-3	Transfer of money from multiple overseas accounts from multiple countries and using the same to buy luxurious items in the range of AU $10,000 to AU $90,000 (vehicles, gold, property).
ML-4	Transfer 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 AU $2,000 to AU $4,900.
ML-5	Deposit a small amount of money in the range of AU $2,000 to AU $4,500 each month to ATM deposit machine and transfer the deposited amount online to an account in a different local bank (but same account) the next day.

ML, machine learning.

the challenges surrounding computing needs, which has helped refine the algorithm further. The automatic feature extraction capability of CNN is another attractive feature for researchers with complex data [44].

This study proposes Conv1D CNN classifier to identify suspicious transactions from financial transaction data. The CNN model captures patterns in the temporal order of the data. Conv1D has fewer parameters than Conv2D, which aids in faster training times and reduces the risk of overfitting. Owing to the limited receptive field of each neuron, the network focuses well on recognizing local patterns in the data, which are beneficial for the detection of suspicious transactions. Conv1D layers help reduce the dimensionality of the data by retaining important features and require fewer computations considering the large-scale nature of finance data. Figure 3 shows the architecture of the proposed CNN classifier, which includes the following layers: Conv1D, normalization, flattening, dropout, and a dense layer. The proposed CNN architecture used five different types of layers from the *Sequential* class, as provided in the Keras library [45], which was built on top of TensorFlow [46]. The layers are described as follows.

Table 2: Overlapping transaction scenarios that shares the characteristics of legitimate and suspicious transactions

S. No.	Scenario description
OL-1	Wire transfer of money to offshore accounts from savings account
OL-2	Cash withdrawal from the account in the range of AU $2,000 to AU $5,000
OL-3	Wire transfer of money from offshore account into savings account
OL-4	Shopping in the range of AU $10,000 to AU $30,000

b.i. Convolution layer

This layer performs the convolution operation on the input data, resulting in an output in the vector format, and then passes it to the next layer [44, 47]. Since financial transactions are time-series data (though not with fixed periods but largely the same pattern), and there is a potential to derive complex features from the data, such as frequency of transactions and volume of

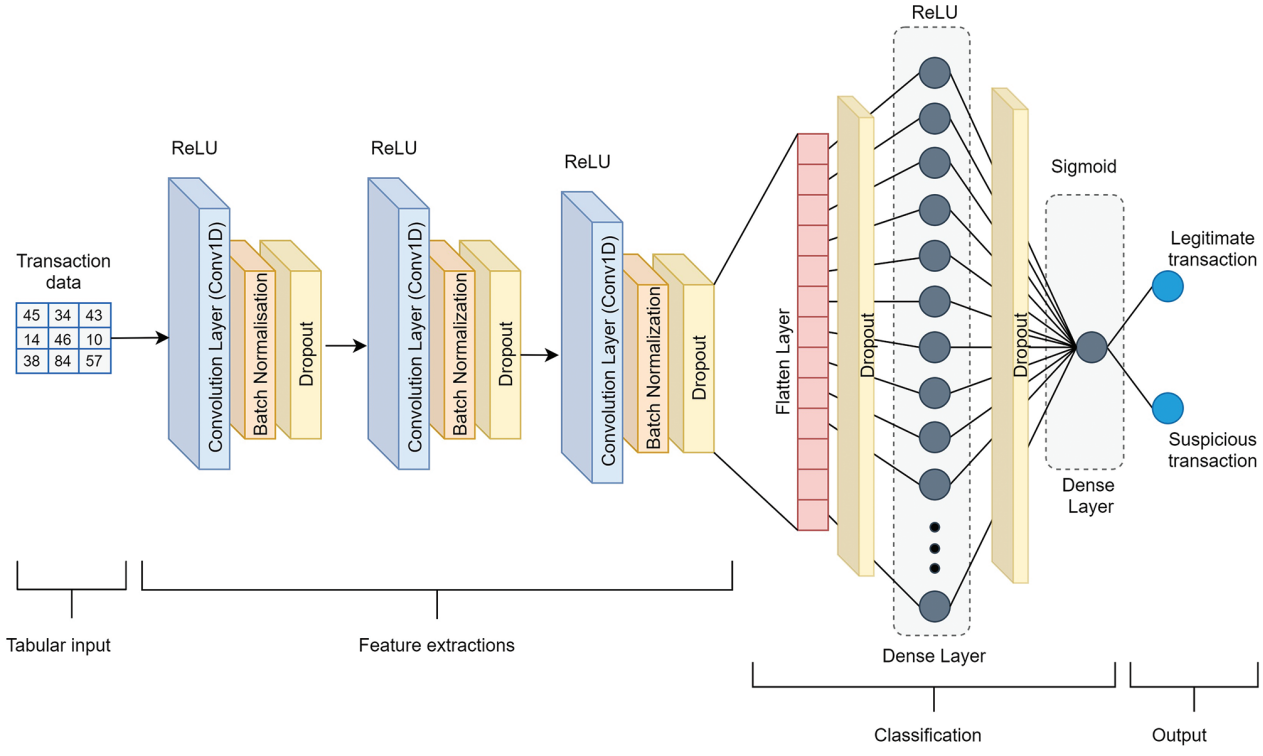


Figure 3: CNN architecture to predict suspicious money laundering transactions. CNN, convolutional neural network.

transactions, we chose the Conv1D type of layer for a tabular dataset to detect suspicious money laundering transactions. The convolution kernel of the Conv1D layer convolves with the layer input over a single dimension to produce the output. A rectified linear unit (ReLU) activation function was used with the Conv1D layer to determine the output. If the input value to the ReLU function is less than zero, it outputs 0; else, it returns the same value as the input. The ReLU activation function also helps prevent exponential growth in the computation required to operate the neural network.

b.ii. Batch normalization layer

This layer applies normalization for input data that helps maintain the output mean close to 0 and a standard deviation close to 1 [48]. During training, the layer normalizes the output by using the mean and standard deviation of the current input batch. During prediction, the layer normalizes the output using the moving average of the mean and standard deviation of the batches observed during training [45].

b.iii. Dropout layer

This layer randomly sets the input to 0 during training at a defined rate frequency, which helps prevent the overfitting issue and improves the model generalization power [45, 49]. The dropout layer is applicable only during training [45].

b.iv. Flatten layer

This layer flattens the input, meaning it converts the multidimensional inputs into a single dimension array that is passed to the dense layer.

b.v. Dense layer

This layer receives input from each neuron of the previous layer and is used to change the dimension of the vector [45, 50]. We used ReLU [51] as an activation function for the hidden dense layer and a sigmoid [52] activation function for the output dense layer.

c. SHAP method for CNN interpretation

At this juncture in AI/ML adoption, when it comes to selecting the classifiers for prediction, there is always a trade-off between accuracy and interpretability. Interpretability refers to how accurately a model associates a cause with an effect. Improved interpretability leads to better model explainability. Explainability refers to the extent to which interpretability can justify predictions. While in some domains such as weather prediction, accuracy may be valued more than model interpretability, in domains where the post-decision stakes are high, model interpretability is valued more to enable adoption. Domains such as medicine [53], regulatory compliance [54, 55] and criminal investigations [56] require the model to be interpretable or explainable. AML is one such domain in which model interpretability is critical [57].

In the banking sector, suspicious transactions detected by the AML system always go through investigation by the AML officer; hence, interpretability can help the officer investigate suspicious transactions more effectively. To reap the benefits of deep learning models, which are black boxes in nature, it is important to understand how such models make decisions. Recently, several *post hoc* methods (such as SHAP [58] and LIME [59]) have been developed to explain the predictions made by opaque or black box models, which apply yet another model on top of the black-box model and tweak the input to see the impact on output values. This helps gain an understanding of the feature importance considered by black-box models for making decisions. Some researchers have a different view that when the decision stakes are high, interpretable methods should be preferred over a *post hoc* approach for interpretability, and future research should focus on creating interpretable models [60].

This study chose a model named agnostic XAI method SHAP [58] to explain the predictions made by the CNN classifier. SHAP is based on the coalitional game theory [61], where feature values are considered team players, the dataset is considered a team, and the game result is considered a prediction. Each feature contributes to the prediction of the results. The SHAP quantifies the contribution of each feature value to the prediction made by the model. 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 predicted target value.

The Shapley value explanation is represented as a linear model and additive feature attribution method [62]. SHAP [58] uses the earlier proposed explanation methods, DeepLIFT [63] and LIME [59], to explain the prediction for the original model $f(z)$, as shown in Eq. (1).

$$f(z) = g(z') = \varnothing_0 + \sum_{j=1}^{M} \varnothing_j Z'_j \quad (1)$$

where $f(z)$ represents the CNN model in this case, z is the input parameter, $g(z')$ represents the explanation model, $z' \in 0, 1$ M is the coalition vector, M is the maximum coalition size, and $\varnothing j \in R$ is a feature j attribution.

SHAP provides multiple explainers such as tree, gradient, deep, and kernel explainers [64]. DeepExplainer, which is an enhanced version of the DeepLIFT [63] algorithm from the SHAP library [64], was used to generate the Shapley values for the CNN classifier and then interpret the predictions using various plots supported by the library.

d. Performance metrics

To measure the performance of classifiers used to detect suspicious transactions, we considered the following metrics.

TP and *true negatives* (TNs) are the transactions that are correctly classified [65, 66]. FPs and FNs are transactions that are incorrectly classified [67, 68]. With respect to the money laundering domain, FPs and FNs have specific importance. A FP indicates that the model has predicted the legitimate transaction as money laundering transaction, and this results in the AML officer manually reviewing it and eventually closing it, which increases the operational cost for the organization. On the contrary, a FN indicates that the model has predicted money laundering transactions as legitimate, which results in missing the actual money laundering transaction, thereby allowing money launderers to use the financial system, which is far more severe and costly than the operational cost of efforts spent on reviewing FPs. More FNs are an indication that the system is not effective and can lead the organization into noncompliance with effective controls, risks of a heavy penalty by regulatory, damage to brand, and corporate citizenship.

Accuracy is the percentage of correctly classified transactions [69] as defined in Eq. (2). Accuracy is one of the best metrics for understanding and measuring model performance, but it is not suitable when the data are highly imbalanced. In financial transaction data and fraud scenarios, the TNs are much higher than the FPs and FNs together, and it undermines the impact of FPs and FNs. Hence, we considered accuracy as a measure to determine the overall model performance, but this was not a key measurement criterion.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{2}$$

Precision is the percentage of predicted positives that are correctly classified [70, 71] as defined in Eq. (3). Usually, if the cost of FPs is higher than that of FNs, then precision is a better metric for measuring the performance of the model. In the money laundering scenario, this is an important metric, as it deals with the operational cost of the effort required for reviewing FPs.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3}$$

Recall is the percentage of actual positives that are correctly classified [71, 72] as defined in Eq. (4). Usually, if the cost of FNs is higher than that of FPs, recall is a better metric. In the money laundering scenario, this is a highly important metric because it involves a high risk of noncompliance, penalties, and brand image due to FNs or missing out on true money laundering transactions.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4}$$

F1 score combines the precision and recall metrics and determines the harmonic mean to measure the performance of the classifier [73] as defined in Eq. (5). For the money laundering detection scenario, to give higher weightage to recall, we have used the following formula to calculate the f_β score with $\beta = 3$.

$$f_\beta \text{score} = \left(1 + \beta^2\right) \times \frac{\text{precision} \times \text{recall}}{\left(\beta^2 \times \text{precision}\right) + \text{recall}} \tag{5}$$

AUC score is used to identify the model's capability to distinguish between legitimate and suspicious transaction classes [74, 75]. The AUC score represents the area under the curve plotted using the TP and FP rates, as defined in Eq. (6). The greater the area under the curve, the better the model.

$$\text{AUC score} = \frac{1}{2}\left(1 + \frac{\text{TP}}{\text{TP} + \text{FN}} - \frac{\text{FP}}{\text{FP} + \text{TN}}\right) \tag{6}$$

IV. Study Results

This chapter presents the synthetic data creation, training, and testing results of CNN model by applying SHAP method on CNN to see the influencing

features on the predictions made by CNN and finally the comparison results of CNN model with RF, SVM, and XGBoost models. Synthetic data were generated using a custom-developed utility by the authors and refined during the experiments. The synthetic dataset contained customers, accounts, legitimate transactions, and suspicious transactions. The code is written to predict suspicious transactions using a CNN classifier. The CNN model was trained and tested on the produced dataset, and its performance was validated against the results produced by RF, XGBoost, and SVM classifiers for the same dataset. Furthermore, the code was written to use the DeepExplainer class of SHAP to produce the Shapley values and explain the predictions made by the CNN classifier at the individual record level. This section provides the results of all the experiments.

a. Synthetic dataset

The key input data for detecting suspicious transactions in banks include customer details, account details, and transactions. Table 3 provides the specifications of the data produced and used for training and testing the models. Transactions are classified into two classes: legitimate and suspicious. By nature, transaction data is highly imbalanced, where the majority class comprises legitimate transactions and the minority class comprises suspicious transactions. In line with this, highly imbalanced data was produced with a ratio of 95.64%–4.36%. The data were validated by an AML expert from the industry.

b. CNN prediction results

The CNN model underwent multiple rounds of training and testing to verify the results by tweaking the hyperparameters. Unlike weights and biases, which are learned by model during the training, the hyperparameters are set prior to training process and are crucial to determine the architecture and behavior of CNN model. We trained and tested the CNN model several times by changing the hyperparameters, including epochs (10–500) that define the number of iterations of passing the entire training data through training process, batch size (16–128) that defines the number of samples used in each iteration of the optimization algorithm, layers (2–6 Conv1D layers; combination of Conv1D, Batch, Dropout, Flatten, and Dense layers) that define the architecture of CNN, dropouts (by dropping and adding after each Conv1D layer)—a regularization technique that randomly drops certain percentage of neurons during training to prevent overfitting, number of units (64–2,048), activation function (ReLU, Softmax, Sigmoid) that introduces nonlinearity in data, allows it to learn complex relationships in the data, and finalizes the parameters. After analyzing various configurations, we finalized the parameters as shown in Table 4, which yielded optimal results with 100 epochs and a batch size of 32. The model was compiled using the Adam optimizer [76] and binary cross entropy loss function [45].

To compare the results of the CNN model, traditional ML models, namely, RF, XGBoost, and SVM, were developed. The hyperparameters of these models are shown in Table 5.

All classifiers considered 70% data for training and 30% data for testing the model. The split data were normally distributed with a feature column mean of 0 and standard deviation of 1 using a standard scaler. This made it easier to apply weights and train the model. Table 6 presents the performance metrics of the predictions made by the CNN, RF, XGBoost, and SVM classifiers. The metrics include confusion matrix (TNs, FNs, FPs, TPs), recall, precision, accuracy, AUC score, f_β score with $\beta = 3$, and training time.

The confusion matrix represents the performance and errors of the classifiers for classification. Type-I errors are shown as FPs, and type-II errors are shown as FNs. The error importance depends on the domain of the classification problem. In the case of AML, we assigned 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. Type-II error indicates that the system is not effective in identifying true suspicious transactions, which is a bigger risk from the compliance point of view and allows the launderers to exploit the financial system for their benefits. Hence, the model should have as few FNs as possible to

Table 3: Synthetic financial transaction dataset summary

Parameter	Value
Number of customers	442
Number of accounts	442
Approximate number of transactions per customer	184
Approximate time period of transactions	12 months
Total number of transactions	92,824
Labeled suspicious transactions	4,054
Labeled legitimate transactions	88,770

Table 4: CNN architecture hyperparameters

Layer	Parameters
Conv1D	Filters = 32, Kernel size = 2, Input shape = 51,980 × 40, Activation = ReLU
Batch normalization	Axis = −1, momentum = 0.99, center = true, scale = true
Dropout	0.3
Conv1D	Filters = 64, Kernel size = 2, Activation = ReLU
Batch normalization	Axis = −1, momentum = 0.99, center = true, scale = true
Dropout	0.3
Conv1D	Filters = 128, Kernel size = 2, Activation = ReLU
Batch normalization	Axis = −1, momentum = 0.99, center = true, scale = true
Dropout	0.3
Flatten	Axis = −1, momentum = 0.99, center = true, scale = true
Dropout	0.3
Dense	Units = 512, Activation = ReLU
Dropout	0.3
Dense	Units = 1, Activation = Sigmoid

CNN, convolutional neural network; ReLU, rectified linear unit.

Table 5: Hyperparameters for RF, XGBoost, and SVM

Classifier	Hyperparameter	Value
RF	Number of trees in the forest	100
RF	Minimum number of data points in a node prior splitting	2
RF	Minimum number of data points allowed in a leaf node	1
RF	Maximum number of features for splitting a node	sqrt
RF	Method for sampling data points	True
RF	Class weight	0:1, 1:100
XGBoost	Minimum number of data points in a node prior splitting	2
XGBoost	Minimum number of data points allowed in a leaf node	1
XGBoost	Learning rate	0.1
XGBoost	Number of decision trees to be boosted	100
XGBoost	Subsample ratio of training data	1
XGBoost	Maximum depth	3
SVM	C	1.0
SVM	Kernel	Linear
SVM	Gamma	Scale

RF, random forest; SVM, support vector machine; XGBoost, extreme gradient boosting.

improve suspicious transaction detection. Type-II error with high importance implies that recall metric is more important than precision; hence, we have considered f_β score metric calculation with $\beta = 3$ to reflect higher importance of recall metric. The results showed that CNN model outperformed other models considering FN identified, recall, and f_β score. The key characteristics of deep learning models are

Table 6: Suspicious transaction prediction results of CNN, RF, XGBoost, and SVM models

Metrics	CNN	RF	XGBoost	SVM
f_β score	78.23%	61.97%	62.09%	30.86%
Recall	91.01%	59.82%	60.14%	29.10%
Precision	34.56%	91.53%	87.72%	67.47%
Accuracy	92.03%	97.95%	97.84%	96.20%
AUC	98.00%	79.80%	98.40%	83.60%
TPs	1,114	746	750	363
TNs	24,515	26,532	26,496	26,426
FPs	2,109	69	105	175
FNs	110	501	497	884
Training time	70 min	6 s	16 s	4.4 min

CNN, convolutional neural network; FNs, false negatives; FPs, false positives; RF, random forest; SVM, support vector machine; TNs, true negatives; TPs, true positives; XGBoost, extreme gradient boosting.

as follows: the accuracy improves with the increase in data size and, in the case of financial domain, millions of transactions are generated every day. This provides a good indication of the application of CNN model in the financial domain.

A FP represents a legitimate transaction reported as a suspicious money laundering transaction by the model. From a domain perspective, each FP transaction alert goes through manual investigation. The impact is that it requires the investigation time of the AML compliance officer, which is widely accepted in banks. Although the aim is to have as less FP as possible, there is a chance of missing out on TPs. Achieving the right balance is crucial, and this balance may come at the cost of the investigation.

FN represents a true money laundering transaction considered by the model as a legitimate transaction, which is a huge risk. The risk lies in the bank's failure to detect the actual money laundering transaction, which results in noncompliance with regulatory policies. If the regulatory authority uncovers such missed transactions later, it may lead to substantial penalties to the bank.

The FN produced by the CNN model is far less than that of other ML models, which indicates that the model can address the risk relatively better than other models. To build trust in the decisions, the XAI model is applied, which shows the rationale behind the decisions, and the compliance officer can quickly take a call by seeing the rationale of the decision to determine whether it is a FP or TP.

From a practical perspective, financial institutions employ several approaches to detect suspicious money laundering transactions. It includes a transaction monitoring system, a rule-based AML system, identifying the topologies by employing different focused approaches, such as detecting anomalies and then investigating from a money laundering perspective, detecting money launderer gangs by detecting patterns using link analysis, graph learning, and social network analysis, and detecting suspicious transactions by applying natural language processing to read through the transaction remarks. AI-based systems often combine multiple classifiers to improve the results. This research presents an approach using one classifier that can be clubbed together while building an AI-based AML system and further enhances the reduction of FPs and FNs. Having FPs is not a "major" concern, as AML officers can still investigate it. FN is riskier as it goes undetected and should be minimized as much as possible. Hence, we believe that this CNN method would help in reducing the number of FNs, and the XAI technique would help the AML Compliance Officer to efficiently investigate the identified suspicious transactions. Figure 4 shows the AUC curves for CNN, RF, XGBoost, and SVM. The AUC measures the ability of the classifier to distinguish between transactions as suspicious or legitimate. The higher the curve, the better the model's ability to distinguish between classification categories.

c. SHAP interpretation of CNN predictions

The models are interpreted at both local and global levels. Local interpretation focuses on determining the reasoning behind the individual prediction, whereas global interpretation focuses on how the model behaves in general. This section describes the local interpretation using the SHAP. After training and testing the CNN classifier, we applied SHAP to interpret the feature importance considered by the CNN while making predictions. A DeepExplainer class from the SHAP library was used to generate the Shapley values for test data containing 24,410 rows and 40 columns, which took approximately 20 min to generate the Shapley values for each data value in the test dataset. The availability of Shapley values for each data element on a record makes it possible to explain the individual record prediction by applying

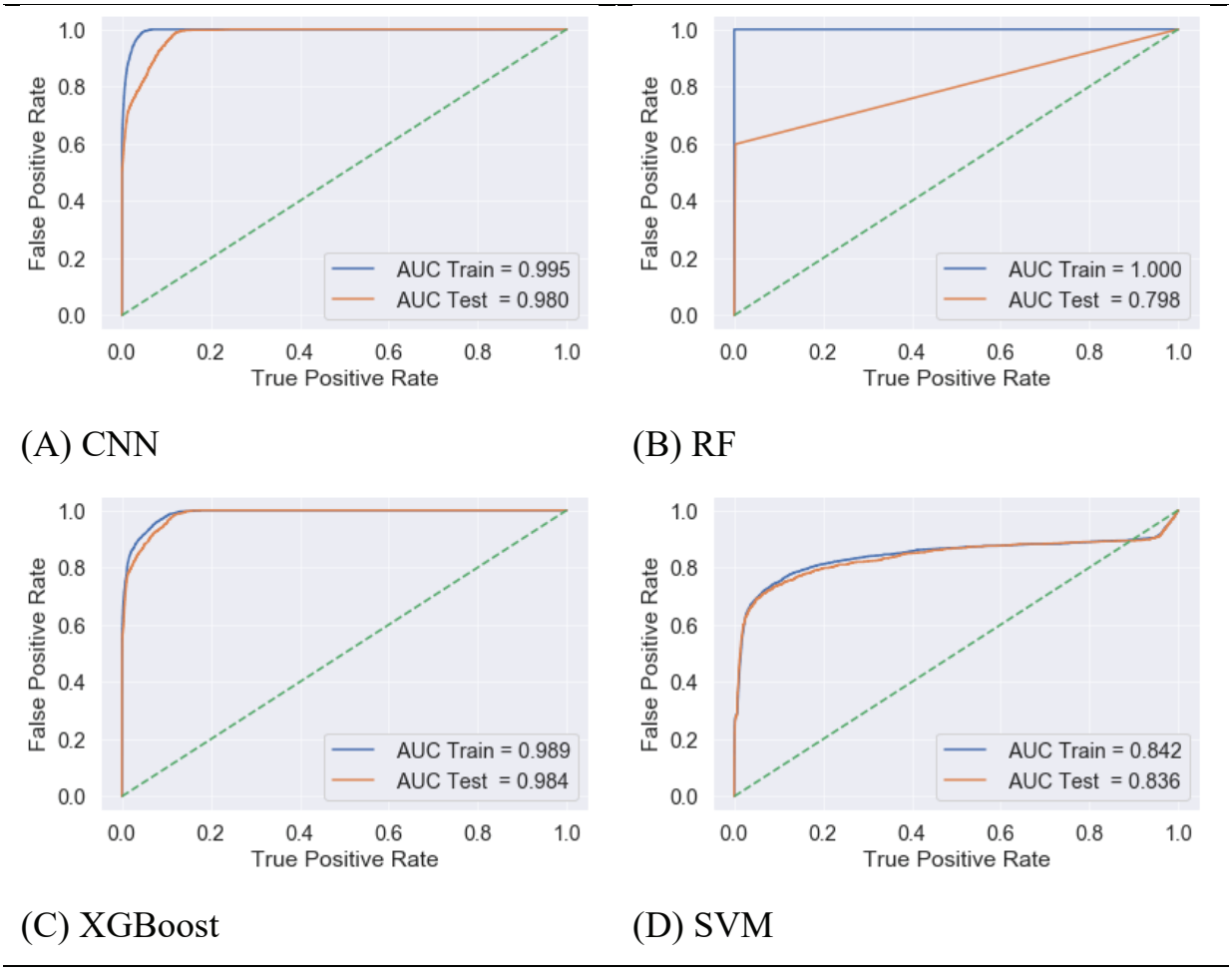
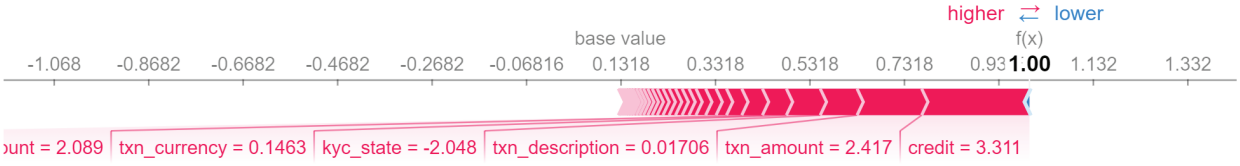


Figure 4: ROC curve for (A) CNN, (B) RF, (C) XGBoost, and (D) SVM classifiers. CNN, convolutional neural network; RF, random forest; SVM, support vector machine; XGBoost, extreme gradient boosting.

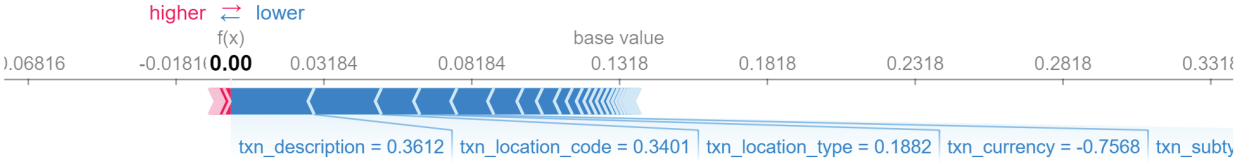
these values to provide a view of the highly contributing features leading to a decision.

Figure 5 shows a force plot representing the interpretation of two individual record predictions made by the CNN classifier. The base value shown in the plot is the average value of the target variable across the dataset that was passed to the DeepExplainer class. Each arrow strip shows the impact of its associated feature on pushing the target variable away from or close to the base value. Red strips show that their associated feature pushes the value on the higher side (indicating a transaction as suspicious) with respect to the base value, whereas the blue strips show that the associated feature pushes the value on the lower side (indicating a transaction as legitimate).

Figure 5A presents the interpretation of one record predicted as a suspicious transaction by the CNN classifier. The key features contributing to the prediction are credit, transaction amount, transaction description, KYC state, and transaction currency. Table 7 shows the details of an original transaction record that is predicted as a suspicious transaction. The table shows the values prior to converting them to numerical and categorical values that are submitted to model. The inference from the force plot is shown in Figure 5A and original transaction record is shown in Table 7. In the "Suspicious Transaction" column, where the customer associated with the transaction account appears to be a student, transaction is done in cash with decent large amount, and the cash is deposited into an account from ATM deposit machine, the account balance appears to be high. This showed a positive correlation with the transaction being suspicious. However, KYC state showed a negative correlation with the prediction because KYC review is complete and status is good; hence,



(A)



(B)

Figure 5: Interpretation of CNN predictions using SHAP force plot. (A) Force plot of a transaction predicted as suspicious by CNN. (B) Force plot of a transaction predicted as legitimate by CNN. CNN, convolutional neural network; SHAP, SHapley Additive exPlanations.

it contributes negatively to the overall prediction. This inference makes the transaction suspicious and worthy of further investigation.

The suspicious transactions reported by financial institutions go through multiple levels of investigation (first by financial institution themselves, then FIUs, investigation agency, and court) before being qualified as a true money laundering transaction. The real value addition by the SHAP explanation for investigation is the identification of key contributing features for a prediction made by the CNN classifier.

Figure 5B represents the interpretation of a record that was predicted as a legitimate transaction by the CNN classifier. The key features that contribute to the prediction are transaction description, transaction location code, transaction location type, transaction currency, and transaction subtype. The "Legitimate Transaction" column in Table 7 shows the details of an original transaction record that is predicted as a legitimate transaction in Figure 5B. The inference from the force plot and original transaction record is that the customer associated with the transaction is a married woman aged 55 years, and the identified transaction is an auto-debit transaction that is paid online for health insurance in the local currency of her birth country, which concludes that the transaction is legitimate. Practically, legitimate transactions in banks will not be screened; however, they are presented here to demonstrate the interpretation to validate the reasoning by SHAP for the prediction made by CNN.

Figure 6 shows the global interpretation or the feature importance graphs for the RF, XGBoost, and SVM classifiers used in comparison with the CNN classifier. Each feature contributed differently to all three models. These are also opaque models, and SHAP can explain individual predictions, which is not considered in the scope of this paper.

We conducted a study to apply the LIME XAI method to the Conv1D model, but this study found that LIME does not support the Conv1D classifier. LIME supports Conv2D. LIME requires data in a three-dimensional format, which is taken as an input by the Conv2D classifier, whereas the input data for Conv1D are two dimensional.

V. Discussion

Banks use rule-based AML systems to generate alerts based on the rules specified by regulators. The alert scenario includes cash transactions beyond thresholds, IFT, and past money laundering topologies. The AML system works as expected for generating alerts for cash transactions and IFT; however, it struggles to detect suspicious transactions. Suspicious transaction alerts are raised by the system based on the scenario developed from the past money laundering patterns. It is difficult to keep these rules up to

Table 7: Sample of two original transaction records considered for prediction by CNN and interpretation by SHAP

Features	Suspicious transaction	Legitimate transaction
Transaction date	7/12/2017	6/02/2018
Transaction number	339549	359932
Transaction account	10300015	10202449
Transaction amount	***6,000.00***	*322.00*
Credit	***6,000.00***	–
Debit	–	*322.00*
Balance	***52,659.00***	16,054.00
Transaction type	Credit	Debit
Transaction subtype	Cash deposit	***Auto-debit***
Transaction description	***Cash deposit***	***Health insurance***
Transaction currency	***AUD***	*AUD*
Transaction location type	***ATM***	***Online***
Transaction location code	448	*222*
Target account	0	891141
Target country code	0	Australia
Target bank code	0	559059
Customer ID	20000736	20002452
Customer 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	Laborers
Income category	4000	77668
KYC updated on date	*22/04/2017*	13/09/2019
KYC state	***Active***	Active
Risk rating	0	0.463290428
Account number	10300015	10202449
BSB number	203901	201807
Account created on date	22/04/2017	23/08/2017
Account type	Savings	Savings
Daily transaction limit	3,000	2,000
TFN	999528645	968305061
Statement delivery method	Not set	Online

BSB, bank state branch; CNN, convolutional neural network; KYC, know your customer; SHAP, SHapley Additive exPlanations; TFN, tax file number.

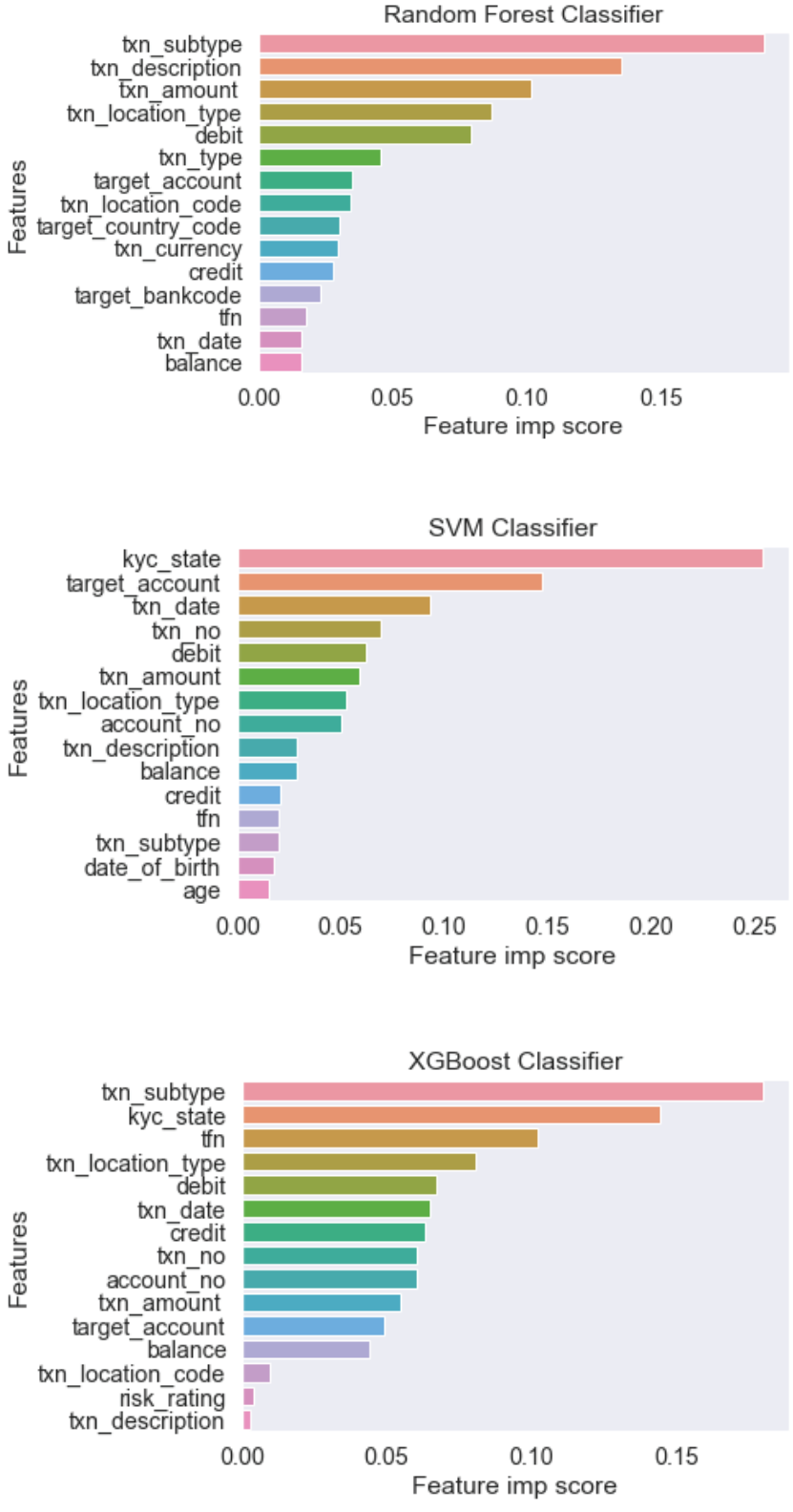


Figure 6: Global interpretation of predictions made by RF, XGBoost, and SVM using feature importance score. RF, random forest; SVM, support vector machine; XGBoost, extreme gradient boosting.

date, considering the pace at which the number of transactions is growing, and new money laundering patterns are being innovated by fraudsters. This causes more FP alerts by the system and impacts the operational efficiency of AML departments involving AML Investigation Officers. AI/ML-based systems that churn out huge amount of data and identify suspicious transactions act as decision support systems for AML officers. These systems are trained using the historical money laundering patterns derived from the alerts generated by the AML system and the details captured from Case and SMR. The key goal of AI/ML-based AML systems is to help AML officers identify suspicious transactions along with adequate evidence. Unidentified money laundering transactions pose a serious risk to banks from a compliance point of view.

The core characteristics of deep learning method to identify features automatically and detect suspicious transactions with high accuracy can help identify suspicious transactions by minimizing FPs and FNs. This is the first study in the money laundering domain that has used a Conv1D CNN model on tabular financial transaction data to detect suspicious money laundering transactions, and successfully trained and tested the Conv1D-based CNN model to classify transactions as either suspicious or legitimate. To compare the CNN results, RF, XGBoost, and SVM classifiers were trained and tested using the same dataset. Furthermore, we applied the state-of-the-art SHAP XAI method to the CNN classifier to interpret decisions at the individual transaction level.

This study generated synthetic data containing customers, accounts, and transactions. By nature, the transaction data used for money laundering detection are highly imbalanced, meaning that legitimate transactions are high in number and suspicious transactions are less compared to legitimate transactions; hence, we followed a similar pattern. The dataset proportion was 95.64%–4.36% for legitimate and suspicious transactions, respectively. Because the data were highly imbalanced, accuracy was not considered a primary metric. We considered f_β score as a key metric of performance measurement with $\beta = 3$ to value recall more than precision. As per the f_β score, we found that that CNN model outperforms RF, XGBoost, and SVM models. This means that the FNs are fewer compared to other models and a good indication from a risk reduction point of view. However, the FP number of the CNN model is higher than that of the other models, which indicates more operational effort for investigation. AML SMEs prioritize reducing risk and ensuring regulatory compliance, over reducing operational efforts. Deep learning methods are believed to perform better with large datasets and, in AML cases, the volume of transactions is growing day by day with digital banking, which makes deep learning an appropriate method to apply.

The interpretation shown by SHAP has identified satisfactory reasoning for the predictions made by CNN, and inspecting those indicators along with the original records containing customer, account, and transaction details provides good insight to AML officers to decide if the identified suspicious transaction is a FP or worth investigating further.

The strength of this study lies in the effective use of the CNN model to identify suspicious transactions with high f_β scores and recall rates. This study used features similar to those used in AML solutions in banks, including customer profiles, accounts, and transactions. We found some customer profile attributes to be key contributors to the detection of suspicious transactions. Furthermore, the research demonstrated that the SHAP XAI technique can be effectively used with the CNN model, providing insights into the feature importance at the individual record level for CNN predictions. SHAP's ability to show the interpretation of CNN predictions by pinpointing the contribution score of each feature to reach a decision at the individual record level would help AML officers investigate suspicious transactions quickly. This will help improve the effectiveness of AML controls at banks and enable the adoption of deep learning in the AML domain.

a. Implications and applications

Banks have established several AML controls to combat money laundering, such as KYC, customer due diligence (CDD), enhanced CDD, risk profiling, watch list screening, sanctions, politically exposed person (PEP) screening, employee training programs, and transaction monitoring for suspicious activities. Considering the continuous increase in transactions, evolving fraud patterns, and changing regulatory requirements, any method that can help identify suspicious transactions is of great value to banks. Primarily, rule-based AML systems (usually bought from established software product companies that are specialized in financial crime) are used by banks for transaction monitoring, which have both advantages and disadvantages. However, the rule-based system is an essential and important system when it comes to detecting transactions for reporting as per the thresholds defined by regulations (e.g., cash transactions and IFTIs). On the positive side,

rule-based systems are configured according to regulatory rules and historical money laundering scenarios, ensuring compliance with reporting thresholds (e.g., IFTIs). However, banks face several challenges in keeping these systems up to date along with the constantly changing regulatory landscape and money laundering patterns. Banks are involved in analyzing AI/ML technology to assist them in detecting suspicious transactions and reducing the compliance risk.

The outcome of this research, a novel method of detecting suspicious transactions using the CNN method, could be highly beneficial for banks. The recommendation for using this model is to ensure that the data attributes and data are in the required format, normalize the data, ingest the actual money laundering transactions as per the historical records available in banks as labeled training data, and train and test the model. It is important to note that the identification of suspicious transactions using CNN should be one of the components in the "to-be" AI-based AML system. The CNN method should be used together with other components to further improve the effectiveness of the outcome, including customer segmentation, risk profiling, social network analysis, customer screening, and sanctions screening, to establish adequate ground to define the suspiciousness of the transactions and report SMRs for compliance. The authors believe that the CNN method can improve the effectiveness of detecting suspicious money laundering transactions, leading to a reduction in compliance risk and operational cost required for the screening and investigation of FP alerts [77].

Furthermore, a bank's future AI-AML system can be enhanced by incorporating the novel method of "explaining CNN predictions using the SHAP XAI framework," as implemented in this research. The explanations generated by *post hoc* SHAP methods for CNN predictions would help AML officers view the reasoning behind the predictions and gather relevant evidence to enhance the investigation and report suspicious transactions with high confidence. An explainable AML system would help build trust [16] and drive the adoption of AI/ML technologies in the AML domain. This adoption can help manage compliance risks, maintain brand reputation, and avoid hefty penalties for noncompliance.

b. Limitations

Banks offer a wide range of products, including accounts, loans, credit cards, insurance, securities, mobile transfers, checks, and money drafts. They serve various types of customers, including retail customers, small businesses, business banking, institutional banking, and investment banking. Additionally, regulatory requirements vary based on the country in which banks operate. Therefore, it is essential to comprehensively test the model while considering all diverse scenarios and data in stages. In this study, the model is tested on a limited set of data, whereas in a real banking scenario there would be millions of transactions per day; hence, the model's performance considering the continuous flood of transactions must be tested. We believe that suspicious transaction detection depends heavily on the availability of historical money laundering transaction data for training.

A limitation of this study is that the models were not trained and tested on real transaction data. All experiments were performed on synthetic data that had limited types of customers, products owned by customers, and types of transactions, whereas in banks there would be a much more complex transaction dataset and access to external data sources such as PEP screening, world checklists, and social network datasets. Hence, the method proposed in this study should be validated on real data before it is considered for actual use.

c. Future research directions

This section illustrates the future research directions in continuation with this study. (1) Deep learning models are primarily designed to work with image type of data; hence, the recommendation is to convert the tabular dataset into images and apply the Conv2D CNN model to check the outcome for suspicious transaction detection. (2) Most state-of-the-art XAI techniques, such as SHAP [58] and LIME [59], support image data to explain the predictions made by deep learning models on image data; hence, converting the tabular data into images and then predicting suspicious transactions using the Conv2D CNN models can open up several XAI options to interpret the predictions. (3) From a data perspective, an optimization study can be performed to eliminate correlated features; for example, the Aquila optimizing technique [78] can be used to reduce the number of features before feeding the CNN model. (4) Attention Mechanism [79] on the CNN model can be used to further reduce FPs. (5) The use of long short-term memory (LSTM) or gated recurrent unit (GRU) model can be explored for predicting suspicious transactions. (6) The research on reinforcement learning methods can be enhanced by leveraging human

decision-making knowledge to detect suspicious transactions.

VI. Conclusions

The aim of this study was to develop a model to predict suspicious money laundering transactions and explain the predictions. The Conv1D CNN classifier was chosen for detecting suspicious transactions and the SHAP XAI method to explain the predictions made by the CNN classifier. Considering the constraint of financial transaction data from banks being highly sensitive and unavailable for research purposes, the authors produced synthetic financial transaction data. The synthetic data was kept highly imbalanced to maintain consistency with real data. In suspicious transaction classification, FPs indicate more operational efforts and FNs indicate a high risk of noncompliance. Considering risk over operational efforts, a recall was given higher weightage over precision for measuring the performance. The results showed that the CNN model successfully identified synthetically injected money laundering transactions far better than other ML models (RF, SVM, and XGBoost) that are used for comparison. The CNN model produced the lowest number of FNs (110), indicating its ability to detect the maximum number of injected suspicious transactions. In comparison, XGBoost produced 497 FNs, RF produced 501 FNs, and SVM produced 884 FNs. This low FN rate is highly favorable for banks, as it reduces the risk of overlooking true money laundering transactions, enhances AML compliance, and helps catch bad actors. Overall, the CNN model outperformed RF, SVM, and XGBoost that is measured using f_β with $\beta = 3$. CNN f_β has the highest score of 78.23 followed by the XGBoost score 62.09, RF score 61.97, and SVM score 30.86. Furthermore, SHAP was successfully applied on CNN to determine the positive or negative contribution of each feature value on the prediction made by CNN. Original transaction records along with SHAP plots were analyzed and found that it was possible to understand the rationale behind the prediction made by CNN, and the same was verified with AML SME. SHAP took around 20 min to generate the Shapley values for 24,410 records, which can be a concern considering millions of records in the bank. In the AML domain, each suspicious transaction alert undergoes manual review and investigation before being reported to regulatory authorities. Therefore, we recommend that the future research should focus on using reinforcement learning methods and leveraging the knowledge gained from alert reviews and investigations as training data to continuously improve model performance.

Author Contribution

Conceptualization: D.K., B.P., and N.S.; data preparation: D.K.; writing the original draft: D.K.; supervision: B.P. and N.S.; methodology: D.K. and B.P.; visualization: D.K., B.P., and N.S.; validation: B.P.; review and editing: B.P. and N.S.; project administration, B.P.; resources: B.P.; funding: B.P. and A.A. All the authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by the Australian Government Research Training Program, the Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), Faculty of Engineering and Information Technology, University of Technology Sydney (UTS), Australia, in part by the Researchers Supporting Project, King Saud University, Riyadh, Saudi Arabia, under Project RSP2024 R14.

References

[1] UNODC, Estimating illicit financial flows resulting from drug trafficking and other transnational organized crimes. 2011, United Nations Office on Drugs and Crime: www.unodc.org.

[2] FATF, *International Standards On Combating Money Laundering And The Financing Of Terrorism & Proliferation - FATF Recommendations*, FATF, Editor. 2023, Financial Action Task Force: https://www.fatf-gafi.org/en/publications/Fatfrecommendations/Fatf-recommendations.html.

[3] BSA, *The Bank Secrecy Act of 1970*, in *12 U.S.C. 1829b, 12 U.S.C. 1951-1960, 31 U.S.C. 5311-5314, 5316-5336*, G.o.U.S.o. America, Editor. 1970, U.S. Government Printing Office: https://www.fincen.gov/resources/statutes-and-regulations/bank-secrecy-act.

[4] AML/CTF, *Anti-Money Laundering and Counter-Terrorism Financing Act 2006*, in *C2023C00383 (C56)*, O.o.P.C. Canberra, Editor. 2006, Government of Australia: https://www.legislation.gov.au/C2006A00169/latest/text.

[5] PMLA, *Prevention of Money Laundering Act, 2002*, in *Act No. 15*, M.o.F. Department of Revenue, Government of India, Editor. 2002, https://enforcementdirectorate.gov.in/pmla.

[6] FinCEN. *Financial Crimes Enforcement Network*. 1990 7 March 2024]; An official website of the United States Government - Financial Crimes Enforcement Network]. Available from: https://www.fincen.gov/.

[7] AUSTRAC. *Australian Transaction Reports and Analysis Centre*. 1989 7 March 2024]; Available from: https://www.austrac.gov.au/.

[8] FIU-India. *Financial Intelligence Unit - India*. 2004 10 March 2024]; Available from: https://fiuindia.gov.in/index.html.

[9] OPA, *Binance and CEO Plead Guilty to Federal Charges in $4B Resolution*, U.S.D.o.J. Office of Public Affairs, Editor. 2023: https://www.justice.gov/opa/pr/binance-and-ceo-plead-guilty-federal-charges-4b-resolution.

[10] Justice, U.S.D.o., *Danske Bank Pleads Guilty to Fraud on U.S. Banks in Multi-Billion Dollar Scheme to Access the U.S. Financial System*, O.o.P.A. Department of Justice, Editor. 2022, United State's Official Government Website: https://www.justice.gov/opa/pr/danske-bank-pleads-guilty-fraud-us-banks-multi-billion-dollar-scheme-access-us-financial.

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

[12] Monroe, B. Fincrime Briefing: AML fines in 2019 breach $8 billion, Treasury official pleads guilty to leaking, 2020 crypto compliance outlook, and more. 2020 7 March 2024]; Available from: 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.

[13] Refinitiv, *Revealing the true cost of financial crime*. 2018: https://www.refinitiv.com/.

[14] McKinsey. *Risk Transforming approaches to AML and financial crime*. 2019 7 March 2024]; Available from: 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.

[15] Al-Shabandar, R., et al., The Application of Artificial Intelligence in Financial Compliance Management, in ACM International Conference Proceeding Series. 2019, ACM. p. 1-6.

[16] PricewaterhouseCoopers. *Explainable AI Driving business value through greater understanding*. 2017 7 March 2024]; Available from: https://www.pwc.co.uk/services/risk-assurance/insights/explainable-ai.html.

[17] PricewaterhouseCoopers, *22nd Annual Global CEO Survey*. 2020: www.pwc.com.

[18] EU, *General Data Protection Regulation (GDPR)*. 2016, Official Journal of the European Union.

[19] State of California, U., TITLE 1.81.5. California Consumer Privacy Act of 2018. 2018.

[20] Kute, D.V., et al., Deep Learning and Explainable Artificial Intelligence Techniques Applied for Detecting Money Laundering-A Critical Review. IEEE Access, 2021. 9: p. 82300-82317.

[21] Kute, D.V., Explainable Deep Learning Approach for Detecting Money Laundering Transactions in Banking System, in Faculty of Engineering and Information Technology. 2022, University of Technology Sydney (UTS), Australia: OPUS. p. 166.

[22] Cutler, A., D.R. Cutler, and J.R. Stevens, *Random forests*, in *Ensemble machine learning*. 2012, Springer. p. 157-175.

[23] Chen, T., et al., *Xgboost: extreme gradient boosting*. R package version 0.4-2, 2015. 1(4): p. 1-4.

[24] Suthaharan, S., Support vector machine, in Machine learning models and algorithms for big data classification. 2016, Springer. p. 207-235.

[25] Chen, Z., et al., Machine learning techniques for anti-money laundering (AML) solutions in suspicious transaction detection: a review. Knowledge and Information Systems, 2018. 57(2): p. 245-285.

[26] Mark Weber, J.C., Toyotaro Suzumura, Aldo Pareja, Tengfei Ma, Hiroki Kanezashi, Tim Kaler, Charles E. Leiserson, Tao B. Schardl, *Scalable Graph Learning for Anti-Money Laundering: A First Look*. 2018.

[27] Alarab, I., S. Prakoonwit, and M.I. Nacer. Competence of graph convolutional networks for anti-money laundering in bitcoin blockchain. In 5th International Conference on Machine Learning Technologies, ICMLT 2020. 2020. Association for Computing Machinery.

[28] Han, J., et al. 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. 2015. Association for Computational Linguistics (ACL).

[29] Paula, E.L., et al. 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. 2017. Institute of Electrical and Electronics Engineers Inc.

[30] Wei, T., et al. A Dynamic Graph Convolutional Network for Anti-money Laundering. In International Conference on Intelligent Computing. 2023. Springer.

[31] Jensen, R.I.T. and A. Iosifidis, *Qualifying and raising anti-money laundering alarms with deep learning.* Expert Systems with Applications, 2023. 214: p. 119037.

[32] Tatulli, M.P., et al. HAMLET: A Transformer Based Approach for Money Laundering Detection. In International Symposium on Cyber Security, Cryptology, and Machine Learning. 2023. Springer.

[33] Silva, Í.D.G., L.H.A. Correia, and E.G. Maziero. Graph Neural Networks Applied to Money Laundering Detection in Intelligent Information Systems. In Proceedings of the XIX Brazilian Symposium on Information Systems. 2023.

[34] Cheng, D., et al., *Anti-Money Laundering by Group-Aware Deep Graph Learning.* IEEE Transactions on Knowledge and Data Engineering, 2023.

[35] Song, J. and Y. Gu, HBTBD: A Heterogeneous Bitcoin Transaction Behavior Dataset for Anti-Money Laundering. Applied Sciences, 2023. 13(15): p. 8766.

[36] Li, Z., et al., Transactional Network Analysis and Money Laundering Behavior Identification of Central Bank Digital Currency of China. Journal of Social Computing, 2022. 3(3): p. 219-230.

[37] Liu, X., X. Wang, and S. Matwin. Interpretable Deep Convolutional Neural Networks via Meta-learning. In 2018 International Joint Conference on Neural Networks, IJCNN 2018. 2018. Institute of Electrical and Electronics Engineers Inc.

[38] Statistics, A.B.o., *Statistics - Australian People and Economy*. 2021, Australian Bureau of Statistics Australian government website.

[39] LeCun, Y., Y. Bengio, and G. Hinton, *Deep learning.* nature, 2015. 521(7553): p. 436-444.

[40] Sabokrou, M., et al., Deep-HR: Fast heart rate estimation from face video under realistic conditions. Expert Systems with Applications, 2021. 186.

[41] Palraj, K. and V. Kalaivani, Predicting the abnormality of brain and compute the cognitive power of human using deep learning techniques using functional magnetic resonance images. Soft Computing, 2021. 25(23): p. 14461-14478.

[42] Li, D., et al., *BLSTM and CNN Stacking Architecture for Speech Emotion Recognition.* Neural Processing Letters, 2021. 53(6): p. 4097-4115.

[43] Roshanzamir, A., H. Aghajan, and M. Soleymani Baghshah, *Transformer-based deep neural network language models for Alzheimer's disease risk assessment from targeted speech.* BMC Medical Informatics and Decision Making, 2021. 21(1).

[44] Miyake, S. and K. Fukushima, A neural network model for the mechanism of feature-extraction - A self-organizing network with feedback inhibition. Biological cybernetics, 1984. 50(5): p. 377-384.

[45] Chollet, F., *Keras: The python deep learning library.* Astrophysics source code library, 2018: p. ascl: 1806.022.

[46] Abadi, M., et al. {TensorFlow}: A System for {Large-Scale} Machine Learning. In 12th USENIX symposium on operating systems design and implementation (OSDI 16). 2016.

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

[48] Ioffe, S. and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In 32nd International Conference on Machine Learning, ICML 2015. 2015. International Machine Learning Society (IMLS).

[49] Srivastava, N., et al., *Dropout: A simple way to prevent neural networks from overfitting.* Journal of Machine Learning Research, 2014. 15: p. 1929-1958.

[50] Huang, G., et al. Densely connected convolutional networks. In 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017. 2017. Institute of Electrical and Electronics Engineers Inc.

[51] Nair, V. and G.E. Hinton. Rectified linear units improve Restricted Boltzmann machines. In 27th International Conference on Machine Learning, ICML 2010. 2010. Haifa.

[52] Han, J. and C. Moraga. The influence of the sigmoid function parameters on the speed of backpropagation learning. In International workshop on artificial neural networks. 1995. Springer.

[53] Tjoa, E. and C. Guan, *A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI.* IEEE Transactions on Neural Networks and Learning Systems, 2021. 32(11): p. 4793-4813.

[54] Huynh, T.D., et al., Addressing Regulatory Requirements on Explanations for Automated Decisions with Provenance—A Case Study. Digital government (New York, N.Y. Online), 2021. 2(2): p. 1-14.

[55] Hall, P., N. Gill, and N. Schmidt, Proposed Guidelines for the Responsible Use of Explainable Machine Learning. 2019.

[56] Hepenstal, S., et al., *Developing Conversational Agents for Use in Criminal Investigations.* ACM transactions on interactive intelligent systems, 2021. 11(3-4): p. 1-35.

[57] Kuiper, O., et al., Exploring Explainable AI in the Financial Sector: Perspectives of Banks and Supervisory Authorities. 2021.

[58] Lundberg, S.M. and S.I. Lee. *A unified approach to interpreting model predictions*. 2017. Neural information processing systems foundation.

[59] Ribeiro, M.T., S. Singh, and C. Guestrin. *"Why should i trust you?" Explaining the predictions of any classifier*. 2016. Association for Computing Machinery.

[60] Rudin, C., Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 2019. 1(5): p. 206-215.

[61] Shapley, L.S., *A Value for N-person Games*. Defense Technical Information Center.

[62] Molnar, C., Interpretable machine learning. A Guide for Making Black Box Models Explainable. 2019, https://christophm.github.io/interpretable-ml-book/.

[63] Shrikumar, A., P. Greenside, and A. Kundaje. Learning important features through propagating activation differences. In 34th International Conference on Machine Learning, ICML 2017. 2017. International Machine Learning Society (IMLS).

[64] Lundberg, S., *SHAP API Library*. 2018: https://shap.readthedocs.io.

[65] Sammut, C. and G.I. Webb, *True Positive*, in *Encyclopedia of Machine Learning*, C. Sammut and G.I. Webb, Editors. 2010, Springer US: Boston, MA. p. 999-999.

[66] Sammut, C. and G.I. Webb, *True Negative*, in *Encyclopedia of Machine Learning*, C. Sammut and G.I. Webb, Editors. 2010, Springer US: Boston, MA. p. 999-999.

[67] Sammut, C. and G.I. Webb, *False Positive*, in *Encyclopedia of Machine Learning*, C. Sammut and G.I. Webb, Editors. 2010, Springer US: Boston, MA. p. 397-397.

[68] Sammut, C. and G.I. Webb, *False Negative*, in *Encyclopedia of Machine Learning*, C. Sammut and G.I. Webb, Editors. 2010, Springer US: Boston, MA. p. 397-397.

[69] Sammut, C. and G.I. Webb, *Accuracy*, in *Encyclopedia of Machine Learning*, C. Sammut and G.I. Webb, Editors. 2010, Springer US: Boston, MA. p. 9-10.

[70] Ting, K.M., *Precision*, in *Encyclopedia of Machine Learning*, C. Sammut and G.I. Webb, Editors. 2010, Springer US: Boston, MA. p. 780-780.

[71] Ting, K.M., *Precision and Recall*, in *Encyclopedia of Machine Learning*, C. Sammut and G.I. Webb, Editors. 2010, Springer US: Boston, MA. p. 781-781.

[72] Sammut, C. and G.I. Webb, *Recall*, in *Encyclopedia of Machine Learning*, C. Sammut and G.I. Webb, Editors. 2010, Springer US: Boston, MA. p. 829-829.

[73] Sammut, C. and G.I. Webb, *F1-Measure*, in *Encyclopedia of Machine Learning*, C. Sammut and G.I. Webb, Editors. 2010, Springer US: Boston, MA. p. 397-397.

[74] Sammut, C. and G.I. Webb, *Area Under Curve*, in *Encyclopedia of Machine Learning*, C. Sammut and G.I. Webb, Editors. 2010, Springer US: Boston, MA. p. 40-40.

[75] Flach, P.A., *ROC Analysis*, in *Encyclopedia of Machine Learning*, C. Sammut and G.I. Webb, Editors. 2010, Springer US: Boston, MA. p. 869-875.

[76] Kingma, D.P. and J. Ba, Adam: A Method for Stochastic Optimization. 2014.

[77] Al-Shabandar, R., et al. The application of artificial intelligence in financial compliance management. In 2019 International Conference on Artificial Intelligence and Advanced Manufacturing, AIAM 2019. 2019. Association for Computing Machinery.

[78] Abualigah, L., et al., *Aquila Optimizer: A novel meta-heuristic optimization algorithm.* Computers & Industrial Engineering, 2021. 157: p. 107250.

[79] Vaswani, A., et al. Attention is All you Need. In Neural Information Processing Systems. 2017.