Money Laundering Detection using Machine Learning and Deep Learning

Johrha Alotibi[1], Badriah Almutanni[2], Tahani Alsubait[3], Hosam Alhakami[4], Abdullah Baz[5]
College of Computers and Information Systems,
Umm Al-Qura University, Makkah, Saudi Arabia[1,2,3,4,5]

***Abstract*—In recent years, money laundering activities have shown rapid progress and have indeed become the main concern for governments and financial institutions all over the world. As per recent statistics, $800 billion to $2 trillion is the estimated value of money laundered annually, in which $5 billion of the total is obtained from cryptocurrency money laundering. As per the financial action task force (FATF), the criminals may trade illegally obtained fiat money for the cryptocurrency. Accordingly, detecting and preventing illegal transactions becomes a serious threat to governments and it has been indeed challenging. To combat money laundering, especially in cryptocurrency, effective techniques for detecting suspicious transactions must be developed since the current preventive efforts are outdated. In fact, deep learning and machine learning techniques may provide novel methods to detect suspect currency movements. This study investigates the applicability of deep learning and machine learning techniques for anti-money laundering in cryptocurrency. The techniques employed in this study are Deep Neural Network (DNN), random forest (RF), K-Nearest Neighbors(KNN), and Naive Bayes (NB) with the bitcoin elliptic dataset. It was observed that the DNN and random forest classifier have achieved the highest accuracy rate with promising findings in decreasing the false positives as compared to the other classifiers. In particular, the random forest classifier outperforms DNN and achieves an F1-score of 0.99%.**

Keywords*—*Anti-money laundering; machine learning; supervised learning; cryptocurrency

I. Introduction

Money laundering is one of the most concerning threats to the stability and progress of the global economy [1]. Such activity is defined as the use of money acquired from illegal activities by hiding the identity of the person and making that money appear legal [2]. Money laundering can also be described as the method of cleaning suspicious money, which represents money collected from illicit or criminal activities, such as illegal gambling, tax evasion, and drug trafficking [3] [4]. Moreover, Integration, layering, and placement are the three primary phases of money laundering. In the placement phase, money is obtained through illicit activities and presented to the system. In the second phase, which is the layers, the source of the funds is hidden through distributing the funds by different intermediaries. In the final phase, the illicit money is transmitted to the criminal [5] [4] [6].

In recent years, money laundering has become more common in news headlines and other forms of media. It has been a fundamentally global problem with social ramifications and economic severe since the mid-1980s [7]. Thus, governments strive to reduce illegal transactions that impact capital [8]. Furthermore, governments around the globe have recommendations and issued regulations for anti-money laundering [3] and are expanding them involving cryptocurrencies [9].

Money laundering has obtained particular attention with the appearance of cryptocurrencies. Bieler illustrates assessed that money laundering earnings are between $ 800 billion to $2 trillion worldwide [10]. About $5 billion of the total is obtained from cryptocurrency money laundering [11].

Because of the anonymity of cryptocurrencies, Campbell-Verduyn [12] discusses that combat money laundering efforts currently require to be improved, because it does not detect money laundering in cryptocurrencies such as Bitcoin, Ethereum, Ripple, and Litecoin. Traditional systems were used by financial institutions specifically on cryptocurrency exchanges, to detect illegal transactions. The results of these traditional systems indicated high low detection rates and high false-positive rates. This means that traditional systems are ineffective at detecting errors and are prone to bias [13]. Traditional systems in financial institutions must be improved and developed in order to detect suspicious transactions [14]. Accordingly, machine learning approaches began being utilized to detect suspicious transactions in 2004 [15]. Thus, studies in recent years have illustrated that the results of using deep learning and machine learning techniques in combating money laundering are indeed promising [16]. Based on machine learning techniques, an anti-money laundering monitoring system is employed at a financial institution in [17] and evaluated using real-life data and feedback from specialized experts. In view of the same, we aim to apply the KNN, NB, RF, and DNN algorithms to the Elliptic Bitcoin dataset to recede the effect of financial crimes on governments and the financial sector. The evaluation models' performance depends on recall, F1-score, and precision, RUC to detect money laundering activities and fraud in cryptocurrency. In addition, it compares the findings with related studies in the same field.

This study is organized as follows. The second section II provides information around the reviews of relevant literature for the study, the reviewed literature covers three major concepts. The third section III presents the research methodology and the deep learning and machine learning techniques used to achieve the results. The fourth section IV presents the details of the data and the preprocessing of the data. The fifth section V presents the final results obtained from the models and compares the previous studies with our results. The final section VI, concludes with a summary of the evaluation of the results obtained and describes the study's limitations and future work.

II. Literature Review

In this section, existing literature related to using of machine learning and deep learning techniques for anti-fraud and money laundering activities in cryptocurrency is reviewed. In fact, cryptocurrencies pose a serious threat to anti-money laundering efforts. In addition, lawbreakers attempt to exploit cryptocurrencies to provide illegal services. As a result, governments such as FATF have developed advanced techniques for anti-money laundering [18]. In view of the same, various techniques and strategies have been widely studied in literature to investigate multiple activities in cryptocurrency data. Essentially, machine learning (ML) and deep learning (DL) are eminent techniques that are capable of investigating vast amounts of data to discover the patterns of illegal financial behavior that have gone undetected [19].

Canhoto [18] and Weber et al. [20], [21] stated that deep learning and machine learning beats the traditional methods of anti-money laundering. Particularly, Weber et al. [21] highlighted the significance of ML regulations and provided the Elliptic dataset for detecting illegal Bitcoin transactions. Different machine learning techniques were used to evaluate the Elliptic dataset, including logistic regression (LR), multilayer perceptrons (MLP), random forest (RF), and graph convolutional networks (GCNs). It was observed that RF technique achieved the high results with a precision, recall-store and F1-score of 0.95, 0.67, and 0.788, respectively. To classify and detect suspicious currency on the Bitcoin network, Lee et al. [22] implemented the artificial neural network (ANN) and RF algorithms. The illegal and legal Bitcoin data were collected from various websites such as Blockchain Explorer and Silk Road. The F1-scores showed that the RF algorithm achieved a high rate of 0.98, while the ANN algorithm achieved a lower rate of 0.89. In the same regard, a novel method for predicting illegal currencies in the Bitcoin currency is proposed by Alarab et al. [23] using a graph convolutional neural network (GCN). The MLP and GCN were combined to enhance the model's performance for which a 0.974 of accuracy was achieved under the proposed method. However, The same author [24] used RF, Extra Trees, Gradient Boosting, XGBoost, LR, and MLP, where RF outperformed with a rate of 0.82. Along similar lines, Ostapowicz and Zbikowski [25] implemented different algorithms on the Ethereum network to identify fraudulent accounts based on supervised learning approach. The accounts were classified and analyzed as "not fraudulen" or "fraudulent" using SVM, XGBoost, and RF. It was observed that the RF algorithm achieved the best results with a detection precision of 85.71. In another study, eight different supervised machine learning techniques were presented and analyzed by Bhowmik et al. [26] to investigate illegal transactions on the blockchain network. These include Naive Bayes (NB), LR, MLP, SVM, RF, Ada Boost, etc. The results of the comparison study found that among the five algorithms, SVM, RF, and NB algorithms obtained the best results with an accuracy of 97%. In view of the same, Monamo et al. [27] also employed an unsupervised learning method based on trimmed k-means and a k-means in order to track down illegal behavior and detect fraudulent activity on the Bitcoin transactions. To classify these transactions, Monamo et al. [28] applied clustering algorithms and machine learning techniques in which several assumptions were imposed to categorize transactions into illegal and legal categories. In addition, different Bitcoin fraud activities were illustrated from both global and local perspectives by using kd-trees and trimmed k-means. To further investigate these two methods, three classification algorithms were used including the maximum likelihood-based, random forests, and boosted binary regression. Based on the obtained results, it was found that the random forest outperformed the other two classification models. Related to the detection and classification of suspected Bitcoin network addresses, several studies have been reported in literature based on different approaches and techniques [13], [29]–[31]. In fact, the unsupervised models for detecting money laundering activities were found to be inadequate for the Bitcoin network as per Lorenz et al. [13]. Therefore, they have developed supervised learning models to identify illegal money laundering activities in the network. In their study, a rule-based technique was employed that showed low detection rates and high false-positive rates. By Lin et al. [29], suspected Bitcoin network addresses transactions were detected and classified by adding the distribution data of transactions, detailed transaction summaries, and time series as new statistics. The model performance was improved and the variance in data was increased. In this study, various machine learning techniques, including LR, SVM, AdaBoost, XGBoost, and LightGBM were implemented. However, LightGBM achieves the best results as compared to the other techniques. A novel method based on a cascade of classifiers and entity characterization to assail bitcoin anonymity was proposed by Zola et al. [30]. In this study, three different algorithms, including the gradient boosting, random forest, and Adaboost, were used to identify illicit transactions on the Bitcoin blockchain network. The inter-entity transactions (organizations or people with multiple accounts) were also investigated, and the classification performance was improved by utilizing 34 features. Bartoletti et al. [31] used data mining and machine learning-based approaches to detect Ponzi schemes related to the Bitcoin addresses. In their study, three machine learning algorithms were provided for evaluation including the Bayes network, random forest, and RIPPER. As a result, the random forest has been proven to detect 96% of addresses. However, it is worth mentioning that the proposed approach was tested against Ponzi schemes.

Kumar et al. [32] classified a 10000-transaction dataset to identify money laundering activities using Naive Bayes algorthoms. The obtained results showed that the proposed model achieved 81% accuracies. In another study, the light gradient boosting machine (LGBM) is proposed by Aziz et al. [33] to detect fraudulent transactions. The MLP, RF, and KNN were compared with the LGBM approach for the identification and classification of fraudulent Ethereum datasets. Relative to the other techniques, the LGBM algorithm has achieved the highest accuracy of 99.03.

Based on the above discussion related to existing literature, it is evident that machine learning algorithms play a vital role in the detection of suspicious transactions in money laundering activities. However, it is worth mentioning that there are still several problems and challenges associated with the detection process that require further improvements. In addition, it seems that there exist very few studies on using deep learning approaches to detect money laundering activities. In view of the same, this paper mainly aims at using deep learning methods with machine learning to detect such suspicious activities in Cryptocurrency.

III. Methodology

The money laundering transaction detection model includes five main stages i.e. data understanding, data preprocessing, data splitting, model training, model testing, and model evaluation. Fig. 1 illustrates the methodological framework of the study. Several ML and DL algorithms are employed in this chapter for transaction classification e.g. NB, RF Classifier, KNN Classifier, and DNN.

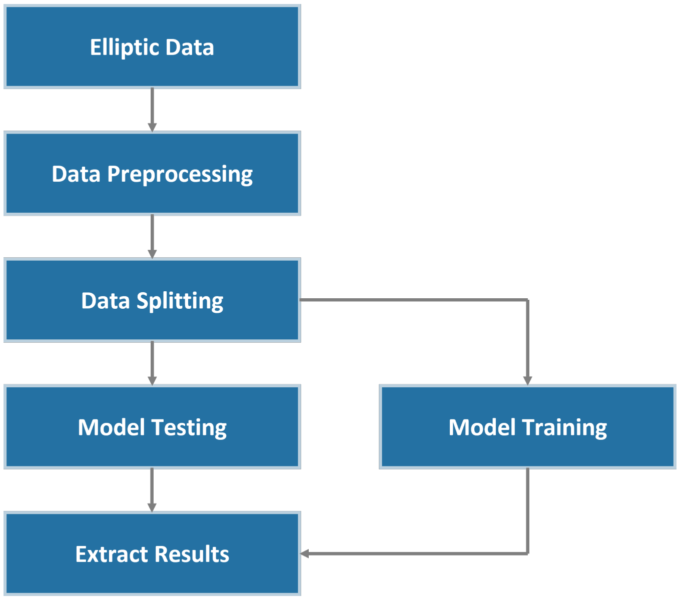

Fig. 1. Methodology.

A. KNN

This approach is considered as a simple instance-based learning algorithm in which the new case/instance is compared with all existing instances and then classified based on a similarity measure. Accordingly, a class is assigned for the new instance based on the nearest available instance. The idea of the instance-based KNN method was first presented by [34]. The measure used is the Minkowski distance.

- Minkowski distance represents the distance between a couple of points in a normalized vector space, and it is defined as following:

$$d(x, y) = (\sum_{i=0}^{n=-1} |x_i - y_i|^{p})^{\frac{1}{p}} \qquad (1)$$

B. RF

Single trees are highly sensitive to training data and might become unstable in certain cases. To overcome this issue, the ensemble strategy is introduced to determine the class label for each data point by enhancing a collection of aggregating and modeling their predictions. On the other hand, decision trees become very popular in data mining due to their simplicity, flexibility, and interpretability especially in handling various data feature types. A RF is represented by a group of regression or classification trees [35]. These groups perform efficiently in case of individual members are not identical.

C. NB

Gaussian Naive Bayes, which uses the Bayes' theorem, is common to the Naive Bayes (NB) algorithms. The Bayesian theorem represents the possibility that an event will happen if you have prior information about a condition associated with the specified event. The method is intended to deal with continuous attributes that are associated with each category and are distributed using a Gaussian distribution. The main advantage of the Naive Bayes is to effectively train in supervised learning, and are used for practical classification problems. A main disadvantage of the Naive Bayes is that the attributes are presumed to be independent, which is nearly impossible to achieve. With Naive Bayes it is considered that all features are independent given the value of a class, this is indicated as conditional independence. There are two categories in this study, illegal transactions = 0 and legal transactions = 1. The Equation 2 shows the likelihood that sample x belongs to a category c

$$P(c|x) = \frac{p(x|c) * p(c)}{P(x)} \qquad (2)$$

D. DNN

DL is a form of ML technique that does not require the construction of feature representation to learn the hierarchical data representation. Instead, it merely uses the training data to automatically learn such representation [36]. This method is based on DNN, which is made up of essential elements including perceptrons, convolutions, and nonlinear activation functions. These elements are structured as layers and trained to understand different complex concepts based on the available raw data. These layers might construct from only a few to over a thousand layers [37]. Lower network layers are typically associated with the low-level features (for example, edges and corners). On the other hand, the higher layers are associated with high-level important features [38].

E. Evaluation Metrics

We use evaluation metrics to evaluate the performance of the model in DL and ML. The evaluation metrics employed for implementing the algorithms are F1-Score, Recall, Precision, and ROC curve. These metrics are commonly applied when dealing with imbalanced datasets, as in the data set used in this study.

Precision refers to the measurement of correct positive predictions in the positive class. The mathematical equation 3 illustrates the concept of precision as follows:

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive} \qquad (3)$$

Recall indicates the number of actual positive data the model was able to correctly predict. The mathematical equation 4 illustrates the concept of recall as follows:

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative} \qquad (4)$$

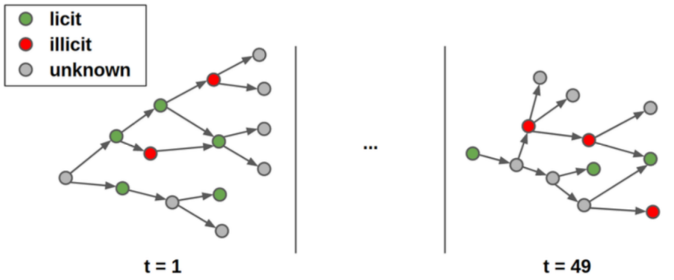

Fig. 2. Structure of the Dataset [39].

F1-score is calculated by the precision and recall value. The mathematical equation 5 illustrates the concept of F1-score as follows:

$$F1 - score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{5}$$

ROC Curve is a graph that measures the performance of a binary classifier across all classification thresholds. The mathematical equations 6 and 7 illustrates the concept of the ROC curve as follows:

$$FPR = \frac{FP}{FP + TN} \tag{6}$$

$$TPR = \frac{TP}{TP + FN} \tag{7}$$

- where FPR stands for False Positive Rate
- where TRP stands for True Positive Rate

IV. Experiments

In this section, a brief overview of the data set used for this study and preprocessing of the data.

A. Dataset

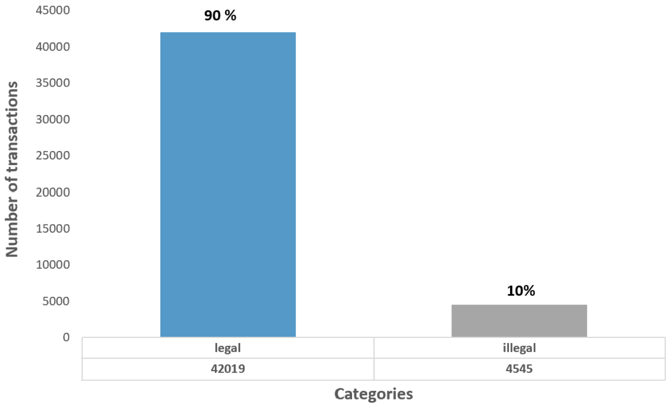

Fig. 3. Distribution of the Transactions.

In this study, the Ellipse dataset[1] created by Weber et al. [21] is employed to detect the cryptocurrency activities. Elliptic is a cryptocurrency monitoring company aimed to protect cryptocurrencies from illegal activity, and it has the largest publicly available dataset for transactions in cryptocurrencies. The dataset contains 49 graphs of BTC transactions obtained at different periods of time. Each graph illustrates a directed acyclic graph (DAG), which means that each edge describes a single directional flow, and there are no loops in the graph. The graph begins from a single transaction and expands to include all the following related transactions, containing two weeks of transaction data created during such period as shown in Fig. 2. As shown in Fig. 2, the BTC transactions are represented in the graph network by nodes (with $203,769$ nodes). On the other hand, the flow of BTC are represented by edges (with $234,355$ edges). The nodes are classified into illegal, legal, and unknown categories. The unknown labels are not considered in this study due to the following reasons:

- The techniques used in this study are based on supervised learning, which requires the ground truth for each data point. Accordingly, supervised learning cannot be used when transactions have unknown labels.
- The number of unknown labels (with 157205 transactions) requires highly efficient hardware resources to train and test models. Unfortunately, such resources are currently not available for the author.

Fig. 3 illustrates the distribution of the elliptic dataset after removing the unknown transactions.

It can be observed that about 10% of the transactions (4,545 samples) are classified as illegal, while 90% (42,019 samples) are classified as legal transactions. Essentially, the legitimate category contains legitimate services, exchanges, and wallet providers, while the illegal category contains scams, Ponzi schemes, terrorist organizations, ransomware, etc. In fact, there exist 166 features associated with each transaction to specify whether they are legal or illegal. Due to the intellectual property rights, the elliptic company has not revealed the nature of the features.

B. Preprocessing of Data

As the model performance can be affected by irrelevant features, it is indeed necessary to detect and select the important features. Particularly, there are 166 features associated with each transaction in the elliptic dataset. Due to intellectual property rights, the elliptic company has not disclosed the details and nature of the features. The class distribution of the dataset is provided in Table I.

TABLE I. Class Distribution of Elliptic Dataset

Label	Number of Samples
Unknown	157,205
legal transactions	42,019
illegal transaction	4,545

It can be observed from Table I that the dataset is unbalanced and contains $157,205$ samples with unknown labels.

[1] Available At: https://www.kaggle.com/datasets/ellipticco/elliptic-data-set

To handle this issue, these unknown samples are eliminated in the first step of data preprocessing. Accordingly, only the samples with legal label ($42,019$ samples) and illegal label ($4,545$ samples) are kept for further steps. Fig. 3 illustrates the distribution of categories after removing the unknown samples. In fact, the dataset contains 166 features which is indeed a large number of features and may consequently lead to overfitting and computational problems. Essentially, the classification techniques require the most relevant features only, which have a high correlation to the class label. In view of the same, a correlation matrix is employed in this study to show the relationship between the features. Accordingly, all features with a correlation greater than 0.90 are eliminated except for the class label. It is worth mentioning that such samples have almost the same effect on the dependent features, and the performance of the model will be significantly affected if no one of them is removed. Based on that, 77 features out of 166 are dropped. However, the number of the remaining features is still large, which is 89 features. Therefore, a preprocessing step has been further implemented to select the most important features based on feature selection techniques through the scikit-learn package. Consequently, the best 53 features have been selected. After choosing the suitable features, we utilize the StandardScaler from scikit-learn for normalizing all features with a standard deviation of 1 and an average value of 0; the purpose is to eliminate bias in classification results.

C. Training and Testing

In this subsection, the model training and testing for transaction classification are discussed. Consequently, the transactions in Elliptic dataset will be classified into legal and illegal transactions. Particularly, the techniques employed in this study are based on supervised learning, which cannot be used when transactions have unknown labels. Therefore, such labels are omitted and not included in the training and testing phases as previously discussed. Essentially, the training set is utilized for model training and hyperparameter tuning. On the other hand, the testing set is utilized to evaluate the performance of the trained model. In the Elliptic dataset, there exist $46,564$ transactions which includes both legal and illegal transactions. The dataset was divided into two parts (70% for training and 30% for training). This is equivalent for $32,594$ transactions for training and $13,969$ transactions for testing including both legal and illegal transactions. The main purpose here is to check and evaluate how the trained model will perform under new transaction data. In fact, a random seed of 42 was determined for splitting the data, which ensures that the data split does not change each time the program is implemented. The data split task was implemented in python through test-train-split from the sklearn library. It is worth noting that two crucial problem associated with machine learning (ML) methods are consequently eliminated under the utilized approach. The first problem is the under-fitting, which is the inability of a ML model to remember the correlations. The second one is the over-fitting, which occurs when a ML algorithm memorizes the patterns.

D. Choice of Algorithms and Hyperparameters

In existing literature, several ML algorithms are used and employed for transaction classification of the elliptic dataset [13], [21], [23], [24], in which the RF algorithm was found to achieve promising results. In [32], [33], NB and KNN algorithms were also used to classify the suspicious transactions and achieved satisfactory results although they were implemented on different data. Based on that, NB, RF, and KNN algorithms are selected in this study in order to achieve high results. On the other hand, DNN techniques are also tested and used in the experiments of this work. It is worth mentioning that DNN techniques have not yet explored and applied on the Elliptic dataset in existing literature. Table II summarizes the hyperparameters utilized in the selected algorithms.

TABLE II. SELECTED ALGORITHM AND HYPERPARAMETERS

Algorithm	Hyperparameters	Description
RF	N-estimatorsint(default=100)	Number of trees
	Max-depth (default=None)	Maximum depth of the tree
	Min_samples_split(default=2)	Minimum of samples
KNN	N-neighbors = 3	Number of neighbors
	weights(default=uniform)	Uniform weights
	Algorithm (default=auto)	Calculate the nearest neighbors
NB	Var-smoothing (default=1e-9)	Portion of the largest variance
DNN	Epoch=10	Total number of iterations
	Optimizer=adam	Adam is an optimization algorithm
	Layer=2	Architecture of the model

In fact, different values are selected during the experiments for hyperparameters of the four models. However, the obtained results were generally unsatisfactory. The results are further improved by choosing the values presented in Table II.

E. Experimental Setup

In this work, the experiments are performed on Core(TM) i7-1065G7 CPU @ 1.30GHz 1.50 GHz based processor, windows 11 with 16.0 GB of RAM. Anaconda environment have been downloaded. In addition, Python 3.7.1 is used as it has a large number of models and libraries available for classification. Some examples of the libraries used in this work include pandas, numpy, seaborn, and matplotlib. The implemented metrics and techniques are obtained by scikit-learn. Tensorflow 2.3.1 and Keras version 2.4.3 are also utilized.

V. RESULTS AND DISCUSSION

The study explores how DL and ML can be used for anti-money laundering using cryptocurrency. This is achieved by using different most common algorithms, including DNN, RF, KNN, and NB, to classify the bitcoin elliptic dataset. discusses the findings obtained in terms of F1 score, recall, precision, and ROC curve and compares our results with previous studies.

A. Results

The machine learning technique has outperformed DNN in classifying legal and illegal transactions. The RF has shown its ability to classify well with an F1 score, precision, ROC curve, and recall of 0.99, the reason for the RF achieving a proper value is the ability to handle an unbalanced dataset. The DNN came in second, which performed an F1 score of 0.98, followed by the KNN with an F1 score of 0.97. When it comes to the NB model, the value is low compared to the RF, KNN, and DNN, it achieved 0.90 in ROC curve and 0.99 in precision. However, the F1-score and recall are only 0.74

and 0.59, respectively. Table III illustrates F1 scores, precision, recall, and ROC curve for each model, using a bold font to highlight the highest value. The RF model has the highest overall value, with an F1 score, precision, recall, and ROC curve. The NB model had the lowest value.

TABLE III. COMPARISON OF THE RESULTS OF THE FOUR MODELS

Models	Measures			
	ROC curve	F1-score	Precision	Recall
NB	0.90	0.74	**0.99**	0.59
RF	**0.99**	**0.99**	**0.99**	**0.99**
KNN	0.92	0.97	0.97	0.98
DNN	0.97	0.98	0.98	0.98

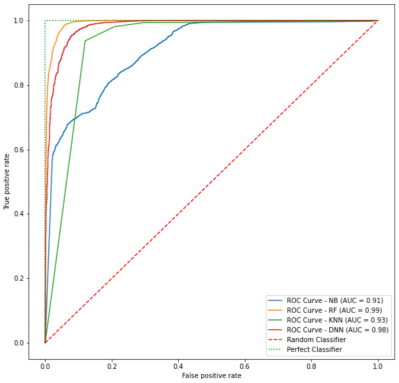

Fig. 4. Comparison of the Results of the Four Models with AUROC Curve

Based on Fig. 4 for the AUROC value for NB, KNN, RF, and DNN, it is shown that the AUC of RF is better than the other three classifiers as it scored in training (1.00) and testing 0.99 while the other models achieved less than 1.00 in training and testing models, but AUROC value in NB model is 0.91, these are the lowest values out of the three models.

B. Discussion and Comparison with Related Works

In comparison to previous studies' results, some studies used DL and ML for the purpose of detecting money laundering in cryptocurrency. Table IV compares the findings of the four previous studies with F1-scores, in [13] which achieved a high value with an f1-score of 0.83 in the RF model, while [23] achieved a value of 0.77% and in [21] achieved a value of 0.78%. Our model outperforms studies by achieving the highest total F1 score of 0.99, as shown in Table IV.

TABLE IV. COMPARISON WITH RELATED WORKS

Ref.	Method	Dataset	Evaluation
Weber et al. [21]	RF - LR - GCNs - MLP	Elliptic dataset	(RF) F1 score =0.78%
Alarab et al. [23]	GCN		(GCN) F1 score =0.77%
Alarab et al. [24]	RF - ExtraTrees -GB - XGBoost - LR -MLP		(RF) F1 score = 0.82%
Lorenz et al. [13]	RF - XGBoost - LR		(RF) F1 score = 0.83%
Current study	**NB-RF -KNN-DNN**		**(RF) F1 score = 0.99%**

VI. CONCLUSION

Money laundering represents a serious threat to governments all over the world and it has been indeed challenging. Various ML and DL techniques have been employed in literature to detect illegal transactions. However, there is still a serious need to further explore and develop suitable algorithms for detecting money-laundering activities, which was the main purpose of the study. Essentially, this research aims to determine the appropriate DL and ML algorithms for detecting money laundering using Elliptic BTC Dataset. To achieve this objective, the results of four algorithms are extensively analyzed and compared. These algorithms include three ML algorithms (RF, KNN, NB), and one DL (DNN). In addition, four key evaluation metrics were used to quantify the performance. These metrics include the precision, recall, F1-score, and ROC curve. the ML technique (RF) proved to be better at classifying fraudulent activities than DL. It was observed from the obtained results that the RF algorithm achieved the best results as compared to other algorithms. It results in 0.99 of the average F1 score. In fact, this technique outperformed the classification due to its ability in handling an unbalanced data set. On the other hand, DNN technique achieved an average F1-score of 0.98 and was placed in the second position followed by the KNN algorithm with an average of 0.97. However, the F1-score for the NB model was found to be 0.74, which is the lowest value as compared to the other three models.

VII. LIMITATIONS AND FUTURE WORK

In fact, the classification model in this study was trained on approximately 46546 bitcoin transactions. However, the dataset contains unlabeled data. To handle this situation, it is more appropriate to use a semi-supervised learning model. However, the unlabeled data will require more CPU power, and therefore, cloud computing services such as Amazon Web Services (AWS) could be used. As the considered model indicates an adequate performance of the algorithms, it would be interesting to conduct the experiment once again with a different data set to prove the validity of the obtained results.

REFERENCES

[1] U. W. Chohan, "The fatf in the global financial architecture: challenges and implications," 2019.

[2] W. Firmansyah and H. T. Atmadja, "Juridical analysis awareness of profession advocacy to financial transaction reports and analysis centre (ppatk) during prevent and eradicate money laundering crime," *Journal of Multidisciplinary Academic*, vol. 5, no. 4, pp. 308–314, 2021.

[3] R. Soltani, U. T. Nguyen, Y. Yang, M. Faghani, A. Yagoub, and A. An, "A new algorithm for money laundering detection based on structural similarity," in *2016 IEEE 7th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*. IEEE, 2016, pp. 1–7.

[4] A. Salehi, M. Ghazanfari, and M. Fathian, "Data mining techniques for anti money laundering," *International Journal of Applied Engineering Research*, vol. 12, no. 20, pp. 10 084–10 094, 2017.

[5] C. Alexandre and J. Balsa, "A multiagent based approach to money laundering detection and prevention." in *ICAART (1)*, 2015, pp. 230–235.

[6] D. Savage, Q. Wang, X. Zhang, P. Chou, and X. Yu, "Detection of money laundering groups: Supervised learning on small networks," in *Workshops at the Thirty-First AAAI Conference on artificial intelligence*, 2017.

[7] G. Sobreira Leite, A. Bessa Albuquerque, and P. Rogerio Pinheiro, “Application of technological solutions in the fight against money laundering—a systematic literature review,” *Applied Sciences*, vol. 9, no. 22, p. 4800, 2019.

[8] S. N. F. S. M. Nazri, S. Zolkaflil, and N. Omar, “Mitigating financial leakages through effective money laundering investigation,” *Managerial Auditing Journal*, 2019.

[9] “Financial Crimes Enforcement Network. 2019. Application of FinCEN’s Regulations to Certain Business Models Involving Convertible Virtual Currencies — FinCEN.gov, howpublished = https://www.fincen.gov/resources/statutes-regulations/guidance/application-fincens-regulations-certain-business-models .”

[10] S. A. Bieler, “Peeking into the house of cards: Money laundering, luxury real estate, and the necessity of data verification for the corporate transparency act’s beneficial ownership registry,” *Fordham J. Corp. & Fin. L.*, vol. 27, p. 193, 2022.

[11] S. Butler, “Criminal use of cryptocurrencies: a great new threat or is cash still king?” *Journal of Cyber Policy*, vol. 4, no. 3, pp. 326–345, 2019.

[12] M. Campbell-Verduyn, “Bitcoin, crypto-coins, and global anti-money laundering governance,” *Crime, Law and Social Change*, vol. 69, no. 2, pp. 283–305, 2018.

[13] J. Lorenz, M. I. Silva, D. Aparício, J. T. Ascensão, and P. Bizarro, “Machine learning methods to detect money laundering in the bitcoin blockchain in the presence of label scarcity,” in *Proceedings of the First ACM International Conference on AI in Finance*, 2020, pp. 1–8.

[14] D. S. Demetis, “Fighting money laundering with technology: A case study of bank x in the uk,” *Decision Support Systems*, vol. 105, pp. 96–107, 2018.

[15] J. McPherson, K.-L. Ma, P. Krystosk, T. Bartoletti, and M. Christensen, “Portvis: a tool for port-based detection of security events,” in *Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security*, 2004, pp. 73–81.

[16] Z. Chen, L. D. Van Khoa, E. N. Teoh, A. Nazir, E. K. Karuppiah, and K. S. Lam, “Machine learning techniques for anti-money laundering (aml) solutions in suspicious transaction detection: a review,” *Knowledge and Information Systems*, vol. 57, no. 2, pp. 245–285, 2018.

[17] P. Tertychnyi, M. Godgildieva, M. Dumas, and M. Ollikainen, “Time-aware and interpretable predictive monitoring system for anti-money laundering,” *Machine Learning with Applications*, vol. 8, p. 100306, 2022.

[18] A. I. Canhoto, “Leveraging machine learning in the global fight against money laundering and terrorism financing: An affordances perspective,” *Journal of business research*, vol. 131, pp. 441–452, 2021.

[19] C. Yan, C. Zhang, Z. Lu, Z. Wang, Y. Liu, and B. Liu, “Blockchain abnormal behavior awareness methods: a survey,” *Cybersecurity*, vol. 5, no. 1, pp. 1–27, 2022.

[20] M. Weber, J. Chen, T. Suzumura, A. Pareja, T. Ma, H. Kanezashi, T. Kaler, C. E. Leiserson, and T. B. Schardl, “Scalable graph learning for anti-money laundering: A first look,” *arXiv preprint arXiv:1812.00076*, 2018.

[21] M. Weber, G. Domeniconi, J. Chen, D. K. I. Weidele, C. Bellei, T. Robinson, and C. E. Leiserson, “Anti-money laundering in bitcoin: Experimenting with graph convolutional networks for financial forensics,” *arXiv preprint arXiv:1908.02591*, 2019.

[22] C. Lee, S. Maharjan, K. Ko, and J. W.-K. Hong, “Toward detecting illegal transactions on bitcoin using machine-learning methods,” in *International Conference on Blockchain and Trustworthy Systems*. Springer, 2019, pp. 520–533.

[23] I. Alarab, S. Prakoonwit, and M. I. Nacer, “Competence of graph convolutional networks for anti-money laundering in bitcoin blockchain,” in *Proceedings of the 2020 5th International Conference on Machine Learning Technologies*, 2020, pp. 23–27.

[24] I. Alarab and S. Prakoonwit, “Effect of data resampling on feature importance in imbalanced blockchain data: Comparison studies of resampling techniques,” *Data Science and Management*, 2022.

[25] M. Ostapowicz and K. Żbikowski, “Detecting fraudulent accounts on blockchain: a supervised approach,” in *International Conference on Web Information Systems Engineering*. Springer, 2020, pp. 18–31.

[26] M. Bhowmik, T. S. S. Chandana, and B. Rudra, “Comparative study of machine learning algorithms for fraud detection in blockchain,” in *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*. IEEE, 2021, pp. 539–541.

[27] P. Monamo, V. Marivate, and B. Twala, “Unsupervised learning for robust bitcoin fraud detection,” in *2016 Information Security for South Africa (ISSA)*. IEEE, 2016, pp. 129–134.

[28] P. M. Monamo, V. Marivate, and B. Twala, “A multifaceted approach to bitcoin fraud detection: Global and local outliers,” in *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2016, pp. 188–194.

[29] Y.-J. Lin, P.-W. Wu, C.-H. Hsu, I.-P. Tu, and S.-w. Liao, “An evaluation of bitcoin address classification based on transaction history summarization,” in *2019 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*. IEEE, 2019, pp. 302–310.

[30] F. Zola, M. Eguimendia, J. L. Bruse, and R. O. Urrutia, “Cascading machine learning to attack bitcoin anonymity,” in *2019 IEEE International Conference on Blockchain (Blockchain)*. IEEE, 2019, pp. 10–17.

[31] M. Bartoletti, B. Pes, and S. Serusi, “Data mining for detecting bitcoin ponzi schemes,” in *2018 Crypto Valley Conference on Blockchain Technology (CVCBT)*. IEEE, 2018, pp. 75–84.

[32] A. Kumar, S. Das, and V. Tyagi, “Anti money laundering detection using naïve bayes classifier,” in *2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON)*. IEEE, 2020, pp. 568–572.

[33] R. M. Aziz, M. F. Baluch, S. Patel, and A. H. Ganie, “Lgbm: a machine learning approach for ethereum fraud detection,” *International Journal of Information Technology*, pp. 1–11, 2022.

[34] D. W. Aha, D. Kibler, and M. K. Albert, “Instance-based learning algorithms,” *Machine learning*, vol. 6, no. 1, pp. 37–66, 1991.

[35] L. Breiman, “Random forests,” *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[36] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[37] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[38] M. D. Zeiler and R. Fergus, “Visualizing and understanding convolutional networks,” in *European conference on computer vision*. Springer, 2014, pp. 818–833.

[39] Bellei, “The elliptic data set: opening up machine learning on the blockchain.” https://medium.com/elliptic/the-elliptic-data-st-opening-up-machine-learning-on-the-blockchain-e0a343d99a14
2010 (accessed Apr 29, 2022).