IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received March 12, 2021, accepted March 31, 2021, date of publication April 9, 2021, date of current version April 26, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3072114

A Time-Frequency Based Suspicious Activity Detection for Anti-Money Laundering

UTKU GÖRKEM KETENCI[1], TOLGA KURT[1], SELİM ÖNAL[2], CENK ERBİL[2], SİNAN AKTÜRKOĞLU[2], AND HANDE ŞERBAN İLHAN[2]

[1]H3M.IO, ITU Teknopark, 34467 Istanbul, Turkey
[2]Akbank T.A.Ş, Sabanci Center, 34330 Istanbul, Turkey

Corresponding author: Utku Görkem Ketenci (utku.ketenci@h3m.io)

ABSTRACT Money laundering is the crucial mechanism utilized by criminals to inject proceeds of crime into the financial system. The primary responsibility of the detection of suspicious activity related to money laundering is with the financial institutions. Most of the current systems in these institutions are rule-based and ineffective (over 90 % false positives). The available data science-based anti-money laundering (AML) models to replace the existing rule-based systems work on customer relationship management (CRM) features and time characteristics of transaction behaviour. Due to thousands of possible account features, customer features, and their combinations, it is challenging to perform feature engineering to achieve reasonable accuracy. Aiming to improve the detection performance of suspicious transaction monitoring systems for AML systems, in this article, we introduce a novel feature set based on time-frequency analysis, that uses 2-D representations of financial transactions. Random forest is utilized as a machine learning method, and simulated annealing is adopted for hyperparameter tuning. The designed algorithm is tested on real banking data, proving the results' efficacy in practically relevant environments. It is shown that the time-frequency characteristics are discriminatory features for suspicious and non-suspicious entities. Therefore, these features substantially improve the area under curve results (over 1%) of the existing data science-based transaction monitoring systems. Using time-frequency features alone, a false positive rate of 14.9% has been achieved, with an F-score of 59.05%. When combined with transaction and CRM features, the false positive rate is 11.85%, and the F-Score is improved to 74.06%.

INDEX TERMS Anomaly detection, anti-money laundering, compliance, random forest algorithm, time-frequency analysis, transaction monitoring.

I. INTRODUCTION

Money laundering (ML) is the umbrella under which the legitimization of the proceeds of crime is attempted while laundered money can be both re-inserted into the legitimate economy and re-used to fuel further criminal activities. All major criminality such as drug and human trafficking, terrorism, extortion, kidnap-for-ransom, bribery, embezzlement, tax evasion, corruption and a multiplicity of other offenses (also known as predicate offenses) are connected through ML. Even though it is impossible to provide an accurate estimate of the size of such a complex underground market, the International Monetary Fund (IMF) indicates that every year, up to 2 trillion USD is laundered through financial systems globally, making ML one of the world's largest markets.

The associate editor coordinating the review of this manuscript and approving it for publication was Keli Xiao.

To tackle this issue, most countries following the Financial Action Task Force (FATF) recommendations set up an anti-money laundering (AML) structure, as shown in Fig.1. It is the responsibility of the financial institutions to report suspicious activities to the Financial Intelligence Unit (FIU). The FIU collects intelligence from all different financial institutions within and outside the jurisdiction, which are later reported to the law enforcement agencies (LEA) as necessary. The police, using this intelligence, builds a case to the judicial system, and if ordered the Asset Recovery Bureau (ARB), recovers the suspicious assets for the public, closing the loop. As the initiator of the whole process, identifying the suspicious activity by the financial institutions is very critical. While technology is essential for the processing and identification of suspicious transactions given the volume of data that needs to be filtered, technology adoption in an AML-context needs to be carefully balanced against the

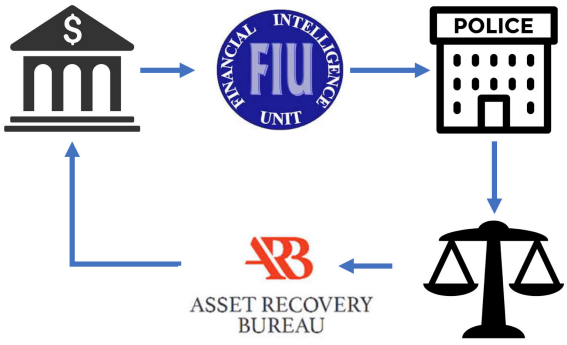

FIGURE 1. The flow of suspicious activity reports through multiple institutions in the judicial system. The detected suspicious activities of bank customers are collected at the FIU, which is reported to the police if necessary. If the judge decides, the ARB acts upon the recovery process of the laundered money. The process in the jurisdiction of Malta is shown as a representative example.

various stakeholders in the AML chain of investigation. Also, most of the existing proposed software' solutions' are rule-based, and only 25 % of respondents have already implemented Artificial Intelligence (AI) and expressed their main business drivers for machine learning in AML as anomaly detection, segmentation, and model tuning [1]. However, the use of AI instead of rule-based engines for new cases detection is infrequent.

The rule-based systems have three significant problems. First, any such software solution depends on a human workforce with varying performance and experience. Instead of enabling AML-analysts and FIU to make more meaningful decisions about what cases should be pursued, they create an unmanageable volume of data. With employees being bombarded by a constant stream of noise from technology-based alerts, it is no surprise that negative repercussions are experienced within financial institutions, which also propagate to the FIUs and the ARBs. As the number of false-positive alerts is over 90% of all alerts, AML experts are consumed by clearing false positives and confirming the non-suspicious nature of cases. This contingency creates difficult work conditions for many AML employees. Often, employees that experience such a continuous stream of false positives will be desensitized towards actual suspicious cases.

Second, most of the suspicious transactions does not even generate and alert since the rules are exposed to criminals from various channels. The exposure can be in the form of insider threats, employees collaborating with money launderers, reverse engineering of software path-dependencies, published Financial Action Task Force (FATF) typologies that are translated into threshold-based rules or contain specific behavioural traits that can be avoided).

Third, the design of rules against new methods of laundering remains a reactive and lengthy process. According to the United Nations Office on Drugs and Crime (UNODC), just 0.2% of the activities can be detected [2]. Despite advances in computation, ML detection remains challenging as a complex behavioural, computational, socio-economic, and managerial problem.

These problems resulted in the introduction of new methods of transaction monitoring using data science and machine learning techniques. However, most of the machine learning techniques are as successful as the quality of the input features. There are hundreds of potential features that can be used, such as ATM withdrawals, SWIFT transactions, online transfers, age, occupation. There are also combinations of features that can be created per channel, per time interval, per currency. The complete list of 237 transactional candidate features (related to the similar field of credit card fraud) has been shown in [3]. As a result, feature engineering (feature creation and selection) for AML is an essential yet very challenging and a time-consuming problem, as specified in [4]–[6]. It can take many weeks, if not months, to determine a useful combination of features out of thousand potential features to be employed.

In this study, we propose a novel and a generalized solution using time-frequency (TF) analysis as a feature extraction method, so that with a handful of features, high-level accuracy can be achieved. Time-frequency features improve the accuracy of machine learning results compared to using transaction features alone. The proposed feature set can be utilized as a standard in suspicious transaction detection in order to shorten the feature engineering stage. There are three key contributions in this work at different stages: feature engineering methodology, model implementation and tests with current real banking data. The first one is a novel methodology for feature extraction in order to build data science models for AML in time and frequency, significantly reducing feature engineering workload. The second contribution is implementing 2D time-frequency features in building data science models for detection, improving the model precision. The third contribution is testing the models in real banking data and proving the improvements in the detection of suspicious activity.

The remainder of the paper proceeds as follows. In the next section, we examine state of the art. In Section III, we present the proposed approach and the time-frequency features. Experiment details and the experimental results are given in Sections IV and V, respectively. Finally, we discuss the results and suggest possible future works.

II. STATE OF THE ART

A. DATA SCIENCE APPROACHES

The suspicious activity detection rules, in essence, try to model the knowledge of the AML subject matter experts. One of the initial surveys of the application of data mining to AML was given in [7]. Many approaches focus on clustering of accounts and transactions and analyzing deviations from clusters and within the cluster for anomalies. For example, Financial Crimes Enforcement Network (FinCEN) has created the FinCEN AI system (FAIS) that links and evaluates reports of large cash transactions to identify potential money laundering; this has been in operation at FinCEN since 1993. The objective is to detect previously unknown, potentially

high-value entities (transactions, subjects, accounts) for possible investigations [8].

In one of the first studies combining domain knowledge in anti-money laundering and data mining, Zdanowicz [9] proposed an approach for outlier detection in under- and over-invoicing. However, global outliers correspond only to a small part of money laundering activities. Therefore, more specific methods have been presented in order to detect local outliers. In [10], outliers have been detected with peer group analysis techniques.

Studies in [11]–[13] applied a cluster-based approach consisting of unsupervised learning techniques such as k-means [14]. Chen *et al.* [15] improved existing outlier detection results with expectation-maximization technique [16]. Unlike the general approach to cluster the customers, Soltani *et al.* [17] cluster the transactions and detect money laundering activities according to structural similarity. In [18], the authors propose a new financial transaction grouping method using a hidden Markov model and genetic algorithm.

Another well-known approach utilized in the field of AML is supervised machine learning. References [19] applied support vector machines to suspicious activity detection. Radial basis function neural network has been used in [20]. On the other hand, the decision tree approach has been applied in [21]–[24]. In another recent work [25], an adaptive neuro-fuzzy inference system is adopted for the AML problem.

Among supervised learning works in the AML domain, [19], [20] have adopted transaction features such as sum and frequency of monetary transactions and [21], [24] have developed models with CRM features. However, there is no example combining these two characteristics.

In terms of results, recall and false positive rate (FPR) have been investigated in [19], [20] and they achieve recall rate between 60% to 80% and false positive rate between 3% to 10%. Recall and FPR depend on the examined data and the selected threshold. In order to minimize the effect of the selected threshold, we examine the area under curve results as an objective evaluation metric in this paper.

The agent-based approaches are proposed in [26] for the detection of suspicious activity by heterogeneous agents called sentinels. In [27], an agent-oriented ontology for monitoring and detecting money laundering process has been presented. In the same way, a multi-agent system architecture has been examined for AML problem in [28]. In [29], the existing multi-agent systems have been extended to combat both fraud and money laundering, and some remarkable results have been presented.

FAIS system had tried data science methods and machine learning. However, due to the low ratio of the number of money laundering samples to all financial transactions, there has been a problem of insufficient labelled data for the machines to learn and train. In order to overcome the problem of labeled sample deficiency, active learning has been proposed [30]. Active learning (AL) is about artificial intelligence asking questions. In order to train the models better and faster, AL identifies samples that require labelling by AML experts so that future machine-based decisions can be re-oriented based on information that is more reflective of the domain and the ontological nature of suspicion.

By representing money transfers as a graph, it is also possible to adopt graph mining methods to AML. Clustering techniques can be used to partition the large graph into the explainable sub-graphs [31]. It is possible to identify the suspicion of the sub-graph using methods such as the ones described in [32]. Frequent sub-graph mining [33] inspired by frequent itemset mining techniques are also applicable to the AML domain. Moreover, a recent study shows the contribution of using social network data over banking transactions data to the accuracy of models [34].

Last but not least, in recent years, blockchain transactions became a significant new area for money laundering. Bitcoin mixer services are investigated [35] and machine learning is adopted to detect wallets used by these services [36]. Ensembling anomaly detection techniques [37], accurate results can be achieved either in fiat currency or bitcoin transaction networks in determining anomalous wallets. Node2vec and random walk can be used to create new feature vectors for nodes, improving the model detection performance [38]. In an another recent study, different type of gradient boosting and random forest algorithms have been applied for the detection of anti-money laundering in cryptocurrency networks [39].

B. TIME-FREQUENCY ANALYSIS

The aforementioned methods utilize two types of features in data science models; Customer Relationship Management (CRM) features and transaction features in time-domain. In this study, we apply a time-frequency spectrogram analysis of transaction data for suspicious activity detection related to anti-money-laundering.

Time-frequency analysis is one of the most potent tools for time-series analysis and has a wide range of applications in multiple domains from security to image processing. The advantages usually arise from the capability of dividing the signals into numerous components in time and frequency for additional signal processing flexibility. Many techniques have been studied in the last years typically differentiate in the way of changing the signal from time to frequency domain [40]. Some of the well-known methods that are utilized are: Fourier transform (FT) consists in decomposing a function into its constituent frequencies, Wigner-Ville distribution, empirical mode decomposition, Gabor transform, and Wavelet transform.

To the best of the authors' knowledge, time-frequency analysis has never been utilized in the context of AML. The idea presented in this work is to transform transaction data into a 2-dimensional time-frequency representation and use statistical features of the time-frequency domain, instead of features of just transactions in the time domain. Time-frequency analysis can be a powerful tool for AML for the following reasons: First, it can provide a complete picture

of characteristics of a single entity and divergence from the peer group; second, it can be representative of behaviour changes from a different perspective. It is expected that the characteristics of routine transactions are much smoother in the time-frequency domain (respectively, suspicious transactions are much sharper) compared to the time domain or frequency domain investigated alone. Thus, it is also expected that the time-frequency domain features will be discriminative for the detection of suspicious transactions.

III. SYSTEM MODEL

A simplified block diagram of the proposed system model is presented in Figure 2. The CRM and transaction database are used for feature creation. Target variable (suspiciousness label) is generated from previously reported suspicious activities. These data sources compose the training set. Machine learning algorithm takes the training set as the input and creates the scores for all the cases. According to a specified threshold, suspicious or clear decision is undertaken. The novelty in the proposed approach is the usage of time-frequency analysis and time-frequency features will be detailed in the following sections.

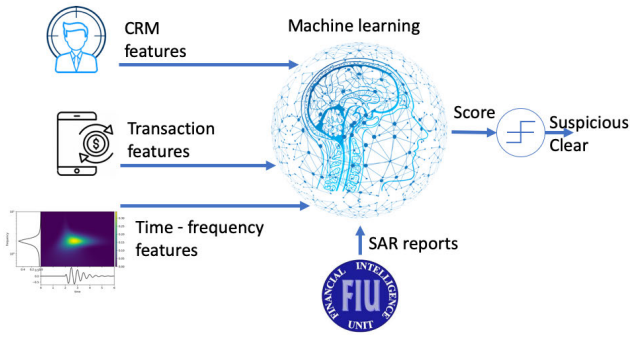

FIGURE 2. System model diagram: The machine learning model (random forest) uses three types of features; CRM features, transaction features, and time-frequency features. The model is trained using previously known suspicious activity. The result of the model is a score between 0 and 1, which is converted into a decision by an optimized threshold. The threshold reflects the risk tolerance of the financial institution.

A. TIME-FREQUENCY REPRESENTATION

To test the effectiveness of the time-frequency features, the following model has been built. For each Akbank customer, the transactions from each banking channel have been modeled as a time series. The funds going in and out to any account of the customer at time t is modeled as $T(n)$. The incoming funds are recorded as positive and outgoing funds as negative.

For a time-series of six months of transactions (signal), $x[n]$, and a time window of w, the discrete-time short Fourier transform (STFT) is defined as

$$STFT(m,w) = \sum_{-\infty}^{+\infty} x[n] \times w[n-m] \times e^{-j\omega n}. \quad (1)$$

Next, the time-frequency domain representation is formed by moving the three month time window one day at a time for the STFT as in Figure 3. For this study, a quarterly sliding time window is utilized, where the granulization of data is daily, i.e. daily total incoming and outgoing funds form the signal. For future work, different granulizations and different sizes of time windows and their combinations can be utilized.

Taking the FT of each window and combining them provides the time-frequency representation of the transactions of a customer, as shown in Figure 4. The figure represents the transactions of a person, who receives salaries on the 15th of every month, pays her rent at the beginning of the month, and spends smaller but random amounts on the remaining days. The repetitive and uniform structure of the representation can be clearly seen in the figure.

Feature extraction in this format is in itself a novel approach, which can be utilized by multiple techniques in future studies. In this work, as a first approach, we focus on features that focus on the energy distribution characteristics as explained below.

B. TIME FREQUENCY FEATURES

The idea proposed in this work is that the time-frequency characteristics of customers that use their accounts for everyday financial transactions will be more natural compared to suspicious people. In order to test this idea, the following 11 metrics of the time-frequency domain representation are calculated:

1) *Mean:* The average value of the time-frequency data points

$$\mu = \frac{1}{T \times 2F} \times \sum_{t=1}^{T} \sum_{f=-F+1}^{F} x(t,f) \quad (2)$$

2) *Variance:* The variance of the time-frequency data points

$$\sigma^2 = \frac{1}{T \times 2F} \times \sum_{t=1}^{T} \sum_{f=-F+1}^{F} [x(t,f) - \mu]^2 \quad (3)$$

3) *Skewness:* A measure of asymmetry in the distribution as an investigation of distance from normal distribution

$$skew = \frac{1}{T \times 2F} \times \sum_{t=1}^{T} \sum_{f=-F+1}^{F} \frac{[x(t,f) - \mu]^3}{\sigma^3} \quad (4)$$

4) *Kurtosis:* The distribution of energy in the FT to the tails of the transformation. In other words, it measures outliers

$$kurt = \frac{1}{T \times 2F} \times \sum_{t=1}^{T} \sum_{f=-F+1}^{F} \frac{[x(t,f) - \mu]^4}{\sigma^4} \quad (5)$$

5) *Time Sparsity:* A measure of sparseness of transactions in the time domain

$$x_t = \sum_{f=-F+1}^{F} x(t,f) \quad (6)$$

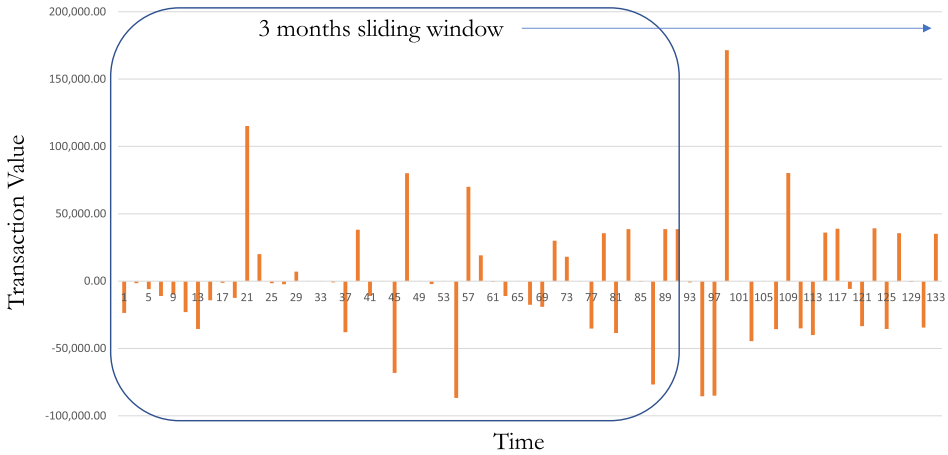

FIGURE 3. Moving time window of transactions (in USD): The bars represent the daily total incoming and outgoing funds, the window of 3 months is slid on daily increments.

$$sparsity_t = \frac{\sum_{t=0}^{T} \begin{cases} 1, & \text{if } x_t \leq \epsilon \\ 0, & \text{otherwise} \end{cases}}{T} \tag{7}$$

6) *Frequency Sparsity:* A measure of sparseness of transactions in the frequency domain

$$x_f = \sum_{t=0}^{T} x(t,f) \tag{8}$$

$$sparsity_f = \frac{\sum_{f=-F+1}^{F} \begin{cases} 1, & \text{if } x_f \leq \epsilon \\ 0, & \text{otherwise} \end{cases}}{2F} \tag{9}$$

7) *Time-Frequency Sparsity:* A measure of sparseness of transactions in the time-frequency domain

$$sparsity_{t,f} = \frac{\sum_{t=0}^{T} \sum_{f=-F+1}^{F} \begin{cases} 1, & \text{if } x_{t,f} \leq \epsilon \\ 0, & \text{otherwise} \end{cases}}{T \times 2F} \tag{10}$$

8) *Time Discontinuity:* The discontinuity of the frequency distribution as the time-window progress

$$disc_t = \sum_{t=0}^{T} \sum_{f=-F+1}^{F-1} x(t,f+1) - x(t,f) \tag{11}$$

9) *Frequency Discontinuity:* The discontinuity of the transaction distribution in consecutive frequencies

$$disc_f = \sum_{f=-F+1}^{F} \sum_{t=0}^{T-1} x(t+1,f) - x(t,f) \tag{12}$$

10) *Time-Frequency Discontinuity:* The discontinuity in the two-dimensional representation

$$disc_f = \sum_{f=-F+1}^{F-1} \sum_{t=0}^{T-1} x(t+1,f) + x(t,f+1) - 2 \times x(t,f) \tag{13}$$

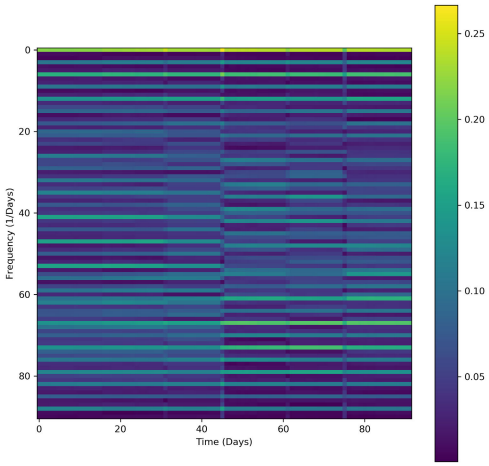

FIGURE 4. Sample normalized time-frequency representation with 90 days of time window: The time-frequency distribution of a normal customer with expected monthly distributions of fund movements.

11) *Entropy:* The dispersion of the information content

$$disc_f = \sum_{f=-F+1}^{F} \sum_{t=0}^{T} -x(t,f) \times \ln(x(t,f)) \tag{14}$$

The hypothesis tested in this work is that, the fundamental energy distribution characteristics of time-frequency representation of suspicious and non-suspicious entities will be different. Using time-frequency features in AML system improvement is the main contribution of this work, whose effectiveness is tested in the next sections.

IV. EXPERIMENTS

A. DATA SET

Unlike most of the studies in this area, instead of using simulated data, the model is tested with real bank data and actual transactions. In order to build the data science model based on various features, Akbank transaction and CRM data for 6.680 customers (who were analyzed in May, 2020) are collected, where among 1.787 of them was related to SAR

(Suspicious Activity Report) activities and 4.893 of them were not deemed suspicious. 6 months of data were analyzed.

B. MODEL TUNING

The whole study has been done on a PC having with 16 GB RAM and i7-8700 3.2 GHz CPU. Random Forest (RF) algorithm [41] with 100 trees is adopted for the sake of performance and accuracy issues. RF is a machine learning technique consisting of an ensemble of decision trees. Three parameters of RF are selected for optimization using simulated annealing (i.e. a metaheuristic technique to approximate the global optimum):

- Minimum numbers of samples to split (*min_split*): the minimum number of samples required to split an internal node in the random forest
- Minimum number of samples to leaf (*min_leaf*): the minimum number of samples required to be at a leaf node (A split point at any depth will only be considered if it leaves at least as many training samples as the parameter value in each of the left and right branches.)
- Maximum depth (*max_depth*): the maximum depth of the tree

To investigate the results, the six different cases are considered as features that are input to the data science model:

1) Training the model with transaction (T) features only
2) Training the model with time-frequency (TF) features only
3) Training the model with customer properties related (CRM) features only
4) Training the model with T + CRM features
5) Training the model with TF + CRM features
6) Training the model with T + TF + CRM

Simulated annealing for 1000 iterations has been run for each feature set for the optimization of min-split, min-leaf and max-depth. Area under the receiver operating characteristic curve (AUC) has been adopted as an objective function to maximize during the optimization. Receiver operating characteristics (ROC) curve represents a true positive rate against false positive rate. AUC corresponds the area under this curve. Higher AUC means higher accuracy. Optimum parameter sets corresponding to each training set are given in Table 1.

TABLE 1. The optimum parameters for different training sets of the random forest algorithm.

Training Set	Optimum min_leaf	Optimum min_split	Optimum max_depth
1	27	30	45
2	8	17	18
3	18	35	54
4	9	10	45
5	4	12	38
6	8	12	36

C. SIZE OF TIME WINDOW

To find approximately an optimum value, a series of trials have been run setting the time window size as 15, 30, 60, 90 and 120 days. AUC values for the six models using only time frequency features are presented with corresponding time window size in Table 2. Among these trials, the best AUC value is achieved with 30 days (1-month) time window.

V. MODEL DETECTION PERFORMANCE

A. COMPARISON OF FEATURE SETS

AUC values for the six models having different feature sets are presented with corresponding time window size in Table 2. It can be seen that only time-frequency features and only transaction features provide almost similar performance (respectively 80.01% and 80.81%). On the other hand, a training set with only CRM features gives a more accurate model with 86.13% AUC. At first glance, we can deduce that traditional CRM features such as age, occupation, etc. are more discriminative.

TABLE 2. AUCs for different sizes of time window.

Size (in day)	15	30	60	90	120	150
AUC (in %)	80.57	80.81	80.34	79.37	77.44	78.56

When time-frequency features are processed with CRM features, the results (90.39% AUC) are approximately 1.4% more accurate in terms of AUC comparing to transaction features with CRM features (giving 88.99% AUC). Thus, we observe that time-frequency features are more appropriate for combining with CRM features.

However, we can note that time-frequency and transaction features have complementary effects and AUC is improved additional 1.1% (91.49% AUC), once all features are combined. Therefore, the AML model becomes more discriminative and more efficient in the detection of suspicious transaction detection.

B. 0.5 THRESHOLD CONFUSION MATRIX

Tables 3,4,5,6,7 and 8 present 0.5 threshold confusion matrices of the models trained respectively with transaction features only, time frequency only, CRM features only, transaction and CRM features, time frequency and CRM features and finally transaction, CRM and time frequency features. In the rows and columns, we present respectively reality (ground truth) and predictions. As aforementioned, predictions are calculated according to a 0.5 threshold, i.e. cases having score greater than 0.5 are considered positive (suspicious). The ideal threshold value may vary from institution to institution depending on the risk tolerance and the amount of workload.

TABLE 3. Confusion Matrix of model trained model with transaction features only.

		Prediction	
		Negative	Positive
Reality	Negative	3518	1375
	Positive	452	1335

We can observe that the number of positive cases predicted as positive (true positive: TP) in Table 3 is 281 more than

TABLE 4. Confusion Matrix of model trained model with time frequency features only.

		Prediction	
		Negative	Positive
Reality	Negative	4164	729
	Positive	733	1054

TABLE 5. Confusion Matrix of model trained model with CRM features only.

		Prediction	
		Negative	Positive
Reality	Negative	3742	1151
	Positive	374	1413

TABLE 6. Confusion Matrix of model trained model with transaction and CRM features.

		Prediction	
		Negative	Positive
Reality	Negative	3975	918
	Positive	353	1434

TABLE 7. Confusion Matrix of model trained model with time frequency and CRM features.

		Prediction	
		Negative	Positive
Reality	Negative	4334	559
	Positive	462	1325

TABLE 8. Confusion Matrix of model trained model with transaction, CRM and time frequency features.

		Prediction	
		Negative	Positive
Reality	Negative	4313	580
	Positive	395	1392

in 4 (and equally the number of negative cases predicted as positive (false positive: FP) in Table 3 is 646 more than in 4). On the other hand, the number of negative cases predicted as negative (true positive: TN) in Table 3 is 646 less than in 4 (and equally the number of positive cases predicted as negative (false positive: FN) in Table 3 is 281 less than in 4). Thus, we can note that the model trained with transaction features are more likely to predict as positive, while the model trained with time frequency features are more likely to predict as negative.

We observe the same effects between Tables 6 and 7 combined with CRM features: The number of TP in Table 6 is more than in 7 and the number of TN in Table 6 is 359 less than in 7.

CRM features minimize the number of FN and give the medium number of FP when comparing with transaction and time frequency features results, as we see in Table 5. Finally, best results have been presented in Table 8 when combining CRM features with both time-frequency and transaction features.

TABLE 9. False Positive Rate, False Negative Rate, Precision, Recall and F-Score comparison according to different training sets.

Training Set	FPR	FNR	PPV	TPR	F-Score
Transaction Features	28.10%	25.30%	49.26%	74.70%	59.37%
Time Frequency Features	14.90%	41.02%	59.11%	58.98%	59.05%
CRM Features	23.52%	20.93%	55.11%	79.07%	64.95%
Transaction and CRM Features	18.76%	19.75%	60.97%	80.25%	69.29%
Time Frequency and CRM Features	11.42%	25.85%	70.33%	74.15%	72.19%
Transaction, CRM and Time Frequency Features	11.85%	22.10%	70.59%	77.90%	74.06%

C. FALSE POSITIVE RATE, FALSE NEGATIVE RATE, PRECISION, RECALL AND F-SCORE

We calculate the false positive rate (FPR), false negative rate (FNR), positive predictive value (PPV or Precision), true positive rate (TPR or Recall) and F-Score for 0.5 threshold as in (15), (16), (17), (18) and (19).

$$FPR = \frac{FP}{FP + TP} \tag{15}$$

$$FNR = \frac{FN}{FN + TN} \tag{16}$$

$$PPV = \frac{TP}{TP + FP} \tag{17}$$

$$TPR = \frac{TP}{TP + FN} \tag{18}$$

$$F\text{-}Score = 2 \cdot \frac{PPV \cdot TPR}{PPV + TPR} \tag{19}$$

The results are presented in Table 9. Our findings are summarized below:

- The best FNR (the minimum value) and the best Recall (the highest value) are achieved with transaction and CRM features. This finding conforms to our assumption (presented in Section V-B) saying that transaction features are more likely to predict cases as positive (and less likely to predict cases as negative). Due to low probability to predict negative, there is also low error in negative cases prediction.
- The best FPR (the minimum value) and the best Precision (the highest value) are achieved with time-frequency and CRM features. This finding is also in line with our assumption (presented in Section V-B) saying that time-frequency features are more likely to predict cases as negative (and less likely to predict cases as positive). Due to the low probability to predict positive, there is also low error in positive cases prediction.
- The best F-Score (the highest value) is achieved with transaction, CRM and time frequency features. In other words, positive effect of combining transaction and time-frequency features is observed in terms of F-Score accuracy for the 0.5 threshold.

D. TEST SET RESULTS

In order to verify generalization of the predictive model; CRM, transaction and time-frequency data for

4.263 customers (who were analyzed in June, 2020 and not analyzed in May, 2020) are collected and examined as test set. 995 of these 4.263 people were related to SAR (Suspicious Activity Report) activities and 3.268 of them were not deemed suspicious. The results are presented in in Fig. 6. There are no major changes compared to the cross-validation results (in Fig. 5). The model with only time-frequency features and only transaction features provide an almost similar performance in cross-validation results and the combination of both features gives again the best AUC score.

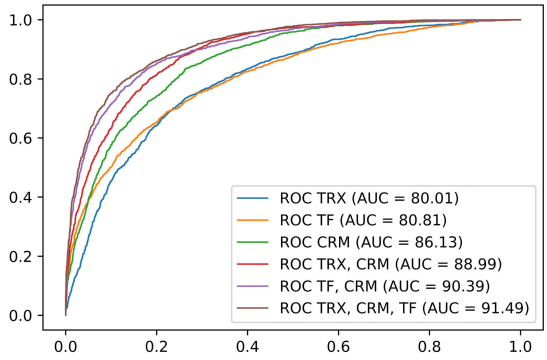

FIGURE 5. ROC curves AUC (in %) comparisons for different models with respect to the involved features (TRX: Transactions, TF: Time-Frequency, CRM: Customer Properties).

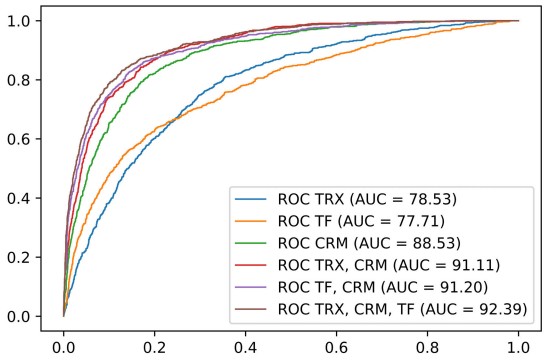

FIGURE 6. ROC curves AUC (in %) comparisons for different models with respect to the involved features in Test Set (TRX: Transactions, TF: Time-Frequency, CRM: Customer Properties).

E. IMPORTANCE OF TIME-FREQUENCY FEATURES

Mutual Information (also known as Information Gain) is generally accepted metric for feature importance in data science solution, first defined in [42]. The mutual information of two jointly continuous random variables X and Y is calculated as

$$MI(X;Y) = \int_Y \int_X p_{X,Y}(x,y) \log\left(\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}\right) dxdy \tag{20}$$

where $p_{X,Y}$ is the joint probability density function of X and Y, and p_X and p_Y are the marginal probability density functions of X and Y respectively.

In our case, X represents a feature in the training set and Y represents its suspiciousness. The relative importance of the features are shown in Table 10. More important features have greater mutual information values.

TABLE 10. Comparison of mutual information of the transaction, CRM and time frequency features.

Rank	Feature	Mutual information (10^{-4})
1	AGE	747.35
2	GENDER	671.17
3	IS-COMMERCIAL	631.07
4	RISK-GROUP	613.54
5	OUTGOING AMOUNT OF FUNDS	612.64
6	OCCUPATION	610.09
7	INCOMING AMOUNT OF FUNDS	566.34
8	CUSTOMER-AGE	341.81
9	KURT	279.49
10	SKEW	276.27
11	ENTROPY	224.76
12	FDISC	171.43
13	MEAN	166.06
14	FSPAR	154.78
15	FTSPAR	149.48
16	TDISC	123.29
17	FTDISC	117.45
18	TSPAR	76.90
19	VAR	17.35

CRM features such as age, gender, boolean showing commercial usage of account, risk group, occupation and number of years of the customer in the bank are among the most important features. We can deduce that the CRM features are more discriminating features. On the other hand, these features can be simply evaluated by analysts at first glance and can be more commonly utilized features by analysts. We can also consider these to be the features that most affect analysts' decisions, due to their simplicity.

In Table 10, we can observe that transaction features such as incoming and outgoing amount of funds are as important as CRM features. However, when combined with CRM features, the contribution of the transaction features to AUC score (presented in Section V) are less than the time frequency features. The reason is the high correlation of incoming and outgoing amount of funds between them (approximately 98 %). Because of the high correlation, these features cannot have an additional effect on the accuracy. On the other hand, time-frequency features are diverse and not all correlated. Therefore, their utilization in the training set is more productive in terms of the AUC improvement.

The analysis shows that smoothness-based features such as Kurtosis and Skewness are more discriminating among the time-frequency features. These features are essential in terms of suspicious activity detection, which is in-line with our assumption regarding the suspicious activity characteristics that the time frequency distributions of normal activities are smooth.

As an example of suspicious activity, in Figure 7, an account behaviour change is shown, where the time-frequency distribution changes significantly and sharply. The change

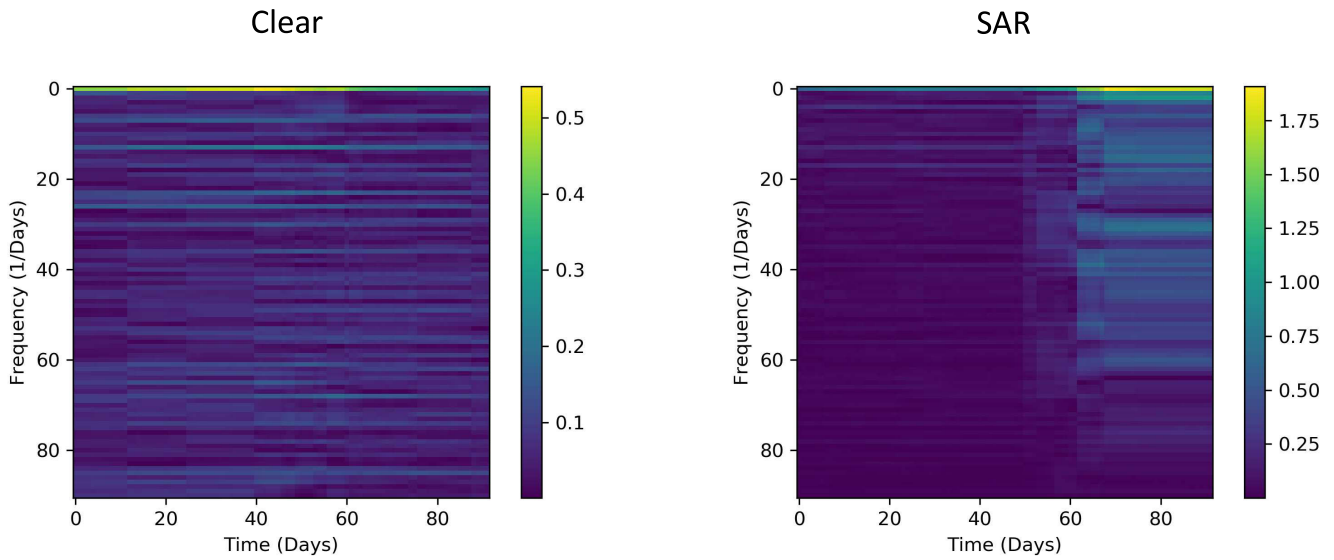

FIGURE 7. Time-frequency plots for sample clear and suspicious cases as a visual for change in account behaviour.

can be seen on the right-hand side of the figure; the frequency characteristics differ significantly for an account whose control has been taken by illicit means.

VI. CONCLUSION AND FUTURE WORK

In this paper, we have shown that adding time-frequency features, simplifies the feature selection process and improves the quality of the data science model. Time-frequency features such as mean, variance, Kurtosis, and skewness have been used for the first time in machine learning model training for suspicious transaction detection. Therefore, the feature engineering stage can be shortened by calculating the proposed time-frequency feature set. This potentially saves many person-months of modeling studies for the financial institutions.

The proposed solution has been implemented in Python and the high-level of accuracy has been proven on real financial data. The generalized solution can easily be adapted to detect suspicious transactions in various organizations. An analysis of actual customer data indicates that time-frequency features can distinguish between suspicious and clear cases, improving AUC and the efficiency of the transaction monitoring system. Among different time-frequency characteristics, Kurtosis provided the maximum differentiation in the model. The gains in accuracy and the capability of detecting money laundering cases that were not detectable before can save financial institutions from regularity fines and HR cost in the order of millions of USD.

In this work, only a low complexity Fourier transform-based approach is utilized for frequency domain analysis. As a future work, the time-frequency analysis can be accomplished with other types of linear and non-linear transforms. There are also potential gains in comparing multiple window lengths, increment sizes and making the analysis in multiple banking channels (such as ATM, Branch, Web). Also, the same analysis can be extended to investigate the characteristics of networks rather than single entities. In particular, when a customer has multiple accounts in multiple banks, the whole picture can only be analyzed by the FIUs. Therefore, repeating this study with additional FIU data would be beneficial as well. Hence, time-frequency features have numerous potential future uses in the area of financial behaviour analysis.

REFERENCES

[1] T. Sausen and A. Liegel, "AI in AML: The shift is underway," NICE Actimize, Hoboken, NJ, USA, Tech. Rep., Jan. 2020. [Online]. Available: https://www.niceactimize.com/Documents/aml_ai_in_aml_insights_report.pdf
[2] *Estimating Illicit Financial Flows Resulting From Drug Trafficking and Other Transnational Organized Crimes*, U. N. O. Drugs and Crime, Vienna, Austria, 2011.
[3] J. Gao, Z. Zhou, J. Ai, B. Xia, and S. Coggeshall, "Predicting credit card transaction fraud using machine learning algorithms," *J. Intell. Learn. Syst. Appl.*, vol. 11, no. 3, pp. 33–63, 2019.
[4] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.
[5] J. Heaton, "An empirical analysis of feature engineering for predictive modeling," in *Proc. SoutheastCon*, Mar. 2016, pp. 1–6.
[6] A. Coates, A. Ng, and H. Lee, "An analysis of single-layer networks in unsupervised feature learning," in *Proc. 14th Int. Conf. Artif. Intell. Statist., JMLR Workshop Conf.*, 2011, pp. 215–223.
[7] R. C. Watkins, K. M. Reynolds, R. Demara, M. Georgiopoulos, A. Gonzalez, and R. Eaglin, "Tracking dirty proceeds: Exploring data mining technologies as tools to investigate money laundering," *Police Pract. Res.*, vol. 4, no. 2, pp. 163–178, Jun. 2003.
[8] T. E. Senator, H. G. Goldberg, J. Wooton, M. A. Cottini, A. U. Khan, C. D. Klinger, W. M. Llamas, M. P. Marrone, and R. W. Wong, "Financial crimes enforcement network AI system (FAIS) identifying potential money laundering from reports of large cash transactions," *AI Mag.*, vol. 16, no. 4, p. 21, 1995.
[9] J. S. Zdanowicz, "Detecting money laundering and terrorist financing via data mining," *Commun. ACM*, vol. 47, no. 5, pp. 53–55, May 2004.
[10] T. Zhu, "An outlier detection model based on cross datasets comparison for financial surveillance," in *Proc. IEEE Asia–Pacific Conf. Services Comput. (APSCC)*, Dec. 2006, pp. 601–604.

[11] Z. Gao, "Application of cluster-based local outlier factor algorithm in anti-money laundering," in *Proc. Int. Conf. Manage. Service Sci.*, Sep. 2009, pp. 1–4.
[12] A. S. Larik and S. Haider, "Clustering based anomalous transaction reporting," *Procedia Comput. Sci.*, vol. 3, pp. 606–610, Jan. 2011.
[13] R. Liu, X.-L. Qian, S. Mao, and S.-Z. Zhu, "Research on anti-money laundering based on core decision tree algorithm," in *Proc. IEEE Chin. Control Decis. Conf. (CCDC)*, May 2011, pp. 4322–4325.
[14] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Statist. Probab.*, vol. 1, Oakland, CA, USA, 1967, pp. 281–297.
[15] Z. Chen, L. Dinh Van Khoa, A. Nazir, E. N. Teoh, and E. K. Karupiah, "Exploration of the effectiveness of expectation maximization algorithm for suspicious transaction detection in anti-money laundering," in *Proc. IEEE Conf. Open Syst. (ICOS)*, Oct. 2014, pp. 145–149.
[16] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc., B, Methodol.*, vol. 39, no. 1, pp. 1–22, 1977.
[17] 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 *Proc. IEEE Annu. Ubiquitous Comput., Electron. Mobile Commun. Conf. (UEMCON)*, Oct. 2016, pp. 1–7.
[18] Y. Li, D. Duan, G. Hu, and Z. Lu, "Discovering hidden group in financial transaction network using hidden Markov model and genetic algorithm," in *Proc. IEEE Int. Conf. Fuzzy Syst. Knowl. Discovery*, vol. 5, Aug. 2009, pp. 253–258.
[19] J. Tang and J. Yin, "Developing an intelligent data discriminating system of anti-money laundering based on SVM," in *Proc. IEEE Int. Conf. Mach. Learn. Cybern.*, vol. 6, Aug. 2005, pp. 3453–3457.
[20] L.-T. Lv, N. Ji, and J.-L. Zhang, "A RBF neural network model for anti-money laundering," in *Proc. IEEE Int. Conf. Wavelet Anal. Pattern Recognit.*, vol. 1, Aug. 2008, pp. 209–215.
[21] S.-N. Wang and J.-G. Yang, "A money laundering risk evaluation method based on decision tree," in *Proc. Int. Conf. Mach. Learn. Cybern.*, vol. 1, Aug. 2007, pp. 283–286.
[22] C. Ju and L. Zheng, "Research on suspicious financial transactions recognition based on privacy-preserving of classification algorithm," in *Proc. IEEE 1st Int. Workshop Educ. Technol. Comput. Sci.*, vol. 2, Mar. 2009, pp. 525–528.
[23] M. A. Villalobos and E. Silva, "A statistical and machine learning model to detect money laundering: An application," Actuarial Sci. Dept. Anahuac Univ., Tech. Rep., 2017. [Online]. Available: https://hddavii.eventos.cimat.mx/sites/hddavii/files/Miguel_Villalobos.pdf
[24] V. Jayasree and R. V. S. Balan, "Money laundering regulatory risk evaluation using bitmap index-based decision tree," *J. Assoc. Arab Universities Basic Appl. Sci.*, vol. 23, no. 1, pp. 96–102, Jun. 2017.
[25] M. B. Jamshidi, M. Gorjiankhanzad, A. Lalbakhsh, and S. Roshani, "A novel multiobjective approach for detecting money laundering with a neuro-fuzzy technique," in *Proc. IEEE 16th Int. Conf. Netw., Sens. Control (ICNSC)*, May 2019, pp. 454–458.
[26] J. Kingdon, "AI fights money laundering," *IEEE Intell. Syst.*, vol. 19, no. 3, pp. 87–89, May 2004.
[27] Y. Wang, H. Wang, S. Gao, D. Xu, and K. Ye, "Agent-oriented ontology for monitoring and detecting money laundering process," in *Proc. Int. ICST Conf. Scalable Inf. Syst.*, 2007, pp. 1–4.
[28] X. Liu and P. Zhang, "An agent based anti-money laundering system architecture for financial supervision," in *Proc. IEEE Int. Conf. Wireless Commun., Netw. Mobile Comput.*, Sep. 2007, pp. 5472–5475.
[29] C. Alexandre and J. Balsa, "A multiagent based approach to money laundering detection and prevention," in *Proc. ICAART*, 2015, pp. 230–235.
[30] X. Deng, V. R. Joseph, A. Sudjianto, and C. F. J. Wu, "Active learning through sequential design, with applications to detection of money laundering," *J. Amer. Stat. Assoc.*, vol. 104, no. 487, pp. 969–981, Sep. 2009.
[31] A. Awasthi, "Clustering algorithms for anti-money laundering using graph theory and social network analysis," Auton. Univ. Barcelona, 2012, p. 75.
[32] C. C. Noble and D. J. Cook, "Graph-based anomaly detection," in *Proc. 9th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2003, pp. 631–636.
[33] S. Aridhi and E. M. Nguifo, "Big graph mining: Frameworks and techniques," *Big Data Res.*, vol. 6, pp. 1–10, Dec. 2016.
[34] A. K. Shaikh, M. Al-Shamli, and A. Nazir, "Designing a relational model to identify relationships between suspicious customers in anti-money laundering (AML) using social network analysis (SNA)," *J. Big Data*, vol. 8, no. 1, pp. 1–22, Dec. 2021.
[35] T. de Balthasar and J. Hernandez-Castro, "An analysis of bitcoin laundry services," in *Proc. Nordic Conf. Secure IT Syst.* Cham, Switzerland: Springer, 2017, pp. 297–312.
[36] J. Wu, J. Liu, W. Chen, H. Huang, Z. Zheng, and Y. Zhang, "Detecting mixing services via mining bitcoin transaction network with hybrid motifs," *IEEE Trans. Syst., Man, Cybern. Syst.*, early access, Jan. 2021, doi: 10.1109/TSMC.2021.3049278.
[37] R. D. Camino, R. State, L. Montero, and P. Valtchev, "Finding suspicious activities in financial transactions and distributed ledgers," in *Proc. IEEE Int. Conf. Data Mining Workshops (ICDMW)*, Nov. 2017, pp. 787–796.
[38] Y. Hu, S. Seneviratne, K. Thilakarathna, K. Fukuda, and A. Seneviratne, "Characterizing and detecting money laundering activities on the bitcoin network," 2019, *arXiv:1912.12060*. [Online]. Available: http://arxiv.org/abs/1912.12060
[39] D. Vassallo, V. Vella, and J. Ellul, "Application of gradient boosting algorithms for anti-money laundering in cryptocurrencies," *Social Netw. Comput. Sci.*, vol. 2, no. 3, pp. 1–15, May 2021.
[40] L. Cohen, *Time-Frequency Analysis*, vol. 778. Upper Saddle River, NJ, USA: Prentice-Hall, 1995.
[41] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[42] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Jul./Oct. 1948.

UTKU GÖRKEM KETENCI received the B.S. degree in computer science from Galatasaray University, Istanbul, in 2008, the M.S. degree in computer science from University Joseph Fourier, Grenoble, in 2009, and the Ph.D. degree in computer science from the University of Valenciennes, Valenciennes, in 2013.

From 2013 to 2014, he was a Research Assistant with the Décision, Interaction, Mobilité Team (DIM), LAMIH, University Polytechnique of Valenciennes. Since 2014, he has been working as a Data Scientist with Banking and Finance Sector. He is the author of several articles in leading journals and conference proceedings. His research interests include multi-agent systems and machine learning.

TOLGA KURT received the B.S. and M.S. degrees in electrical and electronics engineering from Boğaziçi University, Turkey, in 2000 and 2002, respectively, and the Ph.D. degree in electrical engineering from the University of Ottawa, Canada, in 2006.

Following his Ph.D., he was responsible for the 4G product line with Ericsson Canada. He later managed his first startup PlusOneMinusOne working in operational optimisation for banks. He is currently the Managing Partner of AI-Based AML Solutions Company H3M.IO. He has extensive research and development project management experience, which has been translated into productization in more than 40 countries. He has managed more than 15 international software projects, 25 national research and development projects supported by Tubitak, and two EU projects. He has published more than five patents, ten journal articles, and 30 international conference papers. Due to his work in big data analytics, he has been selected to *MIT Technology Review* 35 under 35 list and have been selected as the Young Entrepreneur of the Year in Turkey.

SELİM ÖNAL received the Political Sciences degree from Istanbul University, Istanbul.

From 1991 to 1999, he worked as an Internal Auditor with T.C. Ziraat Bank. From 1999 to 2004, he worked as the Risk Manager with Retail Banking, Koçbank (currently Yapı Kredi Bank). Since 2005, he has been working as the Chief Compliance Officer with the Akbank Compliance Department. He is currently responsible for the Akbank Group Compliance Management, which covers all affiliates in Turkey, Germany, and Malta. He also participates and leads Turkish Banking Association (TBA) Compliance Working Groups, such as the TBA-MASAK (Turkish FIU) Working Group, the National Risk Assesment Group, and the Digital KYC Group. He attended FATF and Private sector meetings in Paris, Brussels, and Amsterdam, as a Turkish Banking Sector Delegate. He also participated as a speaker in many international conferences in London, Paris, Amsterdam, Brussels, Berlin, Istanbul, Ankara, and Antalya.

CENK ERBİL graduated in German from the Business Administrations Department, Marmara University. In 2000, he started his banking career at Koçbank. For 15 out of the 20 years banking experience, he has been working with the Compliance Department, Akbank T.A.Ş. He is currently working as the Financial Crimes Compliance Vice President with Akbank Compliance Department.

SİNAN AKTÜRKOĞLU received the degree from the Faculty of Management, Kocaeli University, in 2002, and the Management M.B.A. degree from Kadir Has University, in 2008. From 2002 to 2004, he worked as an Operation Leader with Martaş Port Companies. From 2004 to 2007, he worked as the Marketing Manager with a family business operating in the textile industry. Since 2007, he has been working with the Compliance Department, Akbank T.A.Ş. He is currently working as a Financial Crimes Investigation and Compliance Manager with the Akbank Compliance Department.

HANDE ŞERBAN İLHAN received the International Relations degree from Istanbul University, Istanbul, in 2014. From 2015 to 2017, she worked as a Financial Auditor with Deloitte. From 2017 to 2018, she worked as a Legislation and Compliance Specialist in a subsidiary with the Garanti BBVA Group. Since 2018, she has been working as a Financial Crimes Investigation and Compliance Specialist with the Akbank Compliance Department.