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

COPY RIGHT

ELSEVIER SSRN

IJIEMR Transactions, online available on 12th September 2024. Link

https://ijiemr.org/downloads.php?vol=Volume-13&issue=issue09

DOI: http://doi.org/10.36893/IJIEMR/2024/V13/I09/8

Title: " A TIME-FREQUENCY BASED SUSPICIOUS ACTIVITY DETECTION FOR ANTI-MONEY LAUNDERING "

Volume 13, ISSUE 09, Pages: 62- 70

Paper Authors
B.Srinivasulu,T. Sravani,T. Ajanta Sravani,T. Vennela

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper as Per UGC Guidelines We Are Providing A ElectronicBarcode

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

A TIME-FREQUENCY BASED SUSPICIOUS ACTIVITY DETECTION FOR ANTI-MONEY LAUNDERING

[1]B.Srinivasulu, [2]T. Sravani,[3]T. Ajanta Sravani,[4]T. Vennela
[1]Assistant Professor,[234]Students
Department Of CSE
Malla Reddy Engineering College for Women

ABSTRACT: The essential tool used by criminals to introduce the profits of their crimes into the financial system is money laundering. Financial institutions have the main obligation for identifying any questionable conduct connected to money laundering. Over 90% of the present systems at these institutions are false positives and rule-based, making them unproductive. The current rule-based systems in use are to be replaced with data science-based anti-money laundering (AML) models that focus on the temporal aspects of transaction behavior and CRM elements. Because there are hundreds of conceivable combinations of customer and account features, it is difficult to do feature engineering with a sufficient level of accuracy. In this paper, we develop a new feature set based on time-frequency analysis to enhance the detection performance of suspicious transaction monitoring systems for AML systems. This feature set leverages 2-D representations of financial transactions. As a machine learning technique, random forest is used, while simulated annealing is used for hyperparameter tweaking. The effectiveness of the developed algorithm is shown in scenarios that are realistically relevant by testing it on actual banking data. It is shown that the time-frequency characteristics may distinguish between things that are suspicious and those that are not. As a consequence, these characteristics significantly raise the current data science-based transaction monitoring systems' area under curve results (above 1%). With an F-score of 59.05%, a false positive rate of 14.9% has been attained using just time-frequency characteristics. The F-Score increases to 74.06% and the false positive rate is reduced to 11.85% when transaction and CRM elements are included.

1. INTRODUCTION:

Under the guise of money laundering (ML), criminals try to normalize the usage of their illicitly obtained funds by either reinvesting them in the regular economy or using them to fund even more illicit endeavors. In ML, all the big crimes—including human and drug trafficking, terrorism, extortion, kidnapping for ransom, bribery, embezzlement, tax evasion, corruption, and a plethora of other crimes—are linked. The International Monetary Fund (IMF) estimates that up to 2 trillion USD is laundered annually via financial institutions internationally, making ML one of the biggest marketplaces in the world. However, it is hard to offer a precise assessment of the scale of such a complicated subterranean sector. As seen in Figure 1, the majority of nations have instituted anti-money laundering (AML) systems in response to this problem, in accordance with the recommendations made by the Financial Action Task Force (FATF). Financial institutions must notify the Financial Intelligence Unit (FIU) of any questionable activity. When required, the FIU will submit the information it has gathered from all financial institutions, both inside and outside of the jurisdiction, to the relevant law enforcement authorities (LEA). Armed with this information, the police file a case with the courts, and if directed to do so, the Asset Recovery Bureau (ARB) retrieves the questionable assets on behalf of the public,

completing the round. Recognizing questionable actions taken by financial institutions is of the utmost importance as the starting point of the whole procedure. Given the sheer amount of data that must be screened, technology is crucial for processing and identifying suspicious activities. However, when it comes to anti-money laundering (AML), it is important to strike a balance between the interests of different stakeholders in the investigative chain. In addition, the majority of the software solutions that have been suggested thus far are rule-based. A mere 25% of respondents have actually put AI into action, and those who have do so have identified anomaly detection, segmentation, and model tuning as the primary business drivers for machine learning in AML. However, novel case identification using AI is still quite rare compared to rule-based engines. The rule-based systems are plagued by three major issues. A human workforce, which might vary in performance and experience, is the first determinant of any software solution of this kind. They generate an overwhelming amount of data instead of helping AML-analysts and FIU make better judgments about which cases to pursue. No one should be surprised that financial institutions, FIUs, and ARBs all feel the effects of the continual barrage of technology-based warnings on staff morale. Since false-positive warnings account for more than 90% of all alerts, anti-money-laundering specialists spend most of their time removing these alerts and verifying that the instances are not suspicious. Many AML personnel find themselves in challenging work situations due to this contingency. Staff members who deal with a high volume of false positives may become immune to real questionable instances. Secondly, considering that the rules are accessible to criminals via several channels, the majority of questionable transactions do not even trigger an alert. Insider threats, staff working with money launderers, software path-dependency reverse engineering, published Financial Action Task Force (FATF) typologies with threshold-based regulations, or avoidable behavioral traits are all examples of what can be exposed. Thirdly, regulations against emerging forms of laundering are yet reactive and time-consuming to draft. Just 0.2 percent of illicit operations are detectable, according to the UNODC [2]. As a multi-faceted issue including behavior, computing, society, and management, ML detection continues to be difficult to solve, even with the advent of new computational tools. Due to these issues, new approaches to transaction monitoring based on data science and machine learning were introduced. The effectiveness of most machine learning methods, however, is directly proportional to the features' quality that are input. Numerous functions are available for usage, including but not limited to: online transfers, age, profession, SWIFT transactions, and ATM withdrawals. Channel-, time-interval-, and currency-specific feature combinations are also possible. In [3], we can see the whole inventory of 237 potential transactional attributes, all of which are associated with the relevant area of credit card fraud. According to [4]-[6], feature engineering, which involves creating and selecting features for AML, is therefore an important yet difficult and time-consuming topic. Finding a practical mix of features from among a thousand possible options might take weeks—if not months—of work. To achieve high-level accuracy with a few of features, we offer a unique and generalized approach in this work that uses time-frequency (TF) analysis as a feature extraction technique. In comparison to relying just on transaction information, machine learning outcomes that include time-frequency features are more accurate. To cut down on feature engineering time, the suggested feature set may be used as a benchmark for suspicious transaction identification. Feature engineering technique, model implementation, and testing using actual

real financial data are the three main contributions of this study, each occurring at a distinct level. Firstly, there is a new approach to feature extraction that may greatly decrease the burden on feature engineers and speed up the process of building data science models for AML. Incorporating 2D time-frequency characteristics into detection data science model construction enhances model accuracy, which is the second contribution. The third addition is that the models have been tested using real-world financial data and the gains in suspicious activity detection have been shown. What follows is the rest of the paper. Here we take a look at the current cutting edge. We lay forth the strategy and time-frequency characteristics in Section III. Sections IV and V include the specifics of the experiment and the outcomes of the experiment, correspondingly. Lastly, we go over the findings and provide some suggestions for potential future research.

1.1 Objective of the project:

The most important tool criminals have for transferring illicitly acquired funds into legitimate financial institutions is money laundering. The onus for identifying questionable actions connected to money laundering mostly rests on financial institutions. Presently, most of these organizations' systems are rule-based, which is both inefficient and inaccurate (false positives exceed 90%). Current anti-money-laundering (AML) methods grounded in data science that aim to replace rule-based systems rely on CRM characteristics and time-related aspects of transaction behavior. It is difficult to execute feature engineering to get tolerable accuracy due to the hundreds of potential account features, customer features, and their combinations. Here, we provide a new set of features based on time-frequency analysis that makes use of two-dimensional representations of financial transactions in an effort to enhance the performance of AML systems' suspicious transaction monitoring systems' detection capabilities. Machine learning is done using random forest, while hyperparameter tweaking is done using simulated annealing. The effectiveness of the findings in realistic settings is shown by testing the created algorithm on actual banking data. The results demonstrate that the time-frequency properties may be used to distinguish between suspicious and non-suspicious things. Consequently, the area under curve results of the current data science-based transaction monitoring systems are significantly improved (over 1%), thanks to these capabilities. A false positive rate of 14.9% and an F-score of 59.05% were attained using time-frequency characteristics alone. Using transaction and CRM features together improves the F-Score to 74.06% and reduces the false positive rate to 11.85%.

2. LITERATURE SURVEY:

"Analyzing the Illicit Funds Derived from International Narcotics Trade and Related Offenses"

According to this study's analysis of the available literature, transnational organized crime laundered almost 2.7% of the world's GDP in 2009, or $1.6 trillion USD, after receiving it via illegal means. Twenty percent of all criminal revenues seem to originate from the sale of illegal narcotics, making it the primary source of cash for global organized crime. Compared to the about $1 billion given to farmers in the Andean area, this analysis estimates that worldwide cocaine sales in 2009 generated gross profits of $84 billion. North America and West and Central Europe accounted for the lion's share of the retail and wholesale gross earnings, with $35 billion and $26 billion, respectively. The paper serves as a reminder to readers that the "crowding out" of legal economic sectors and inefficiencies in

resource allocation are among the many difficulties that might arise when illicit funds are invested in licit economies. Nations in Europe, South America, and North America account for the lion's share of the money leaving the country in order to be laundered. The sum of all money that leaves these areas from cocaine sales is 95 percent. According to the research model, the Caribbean would be the primary location outside of the profit-generating areas for net outflows (outflows minus inflows), with net inflows of around US $6 billion. "Predicting credit card transaction fraud using machine learning algorithms," a voluminous body of data accompanied by accompanying language from relevant international law instruments, (

Theft and fraudulent use of payment cards constitute credit card fraud, a pervasive problem for banks and other financial organizations. In this research, we investigate how machine learning models, together with linear and nonlinear statistical models, might be applied to actual data obtained from credit card transactions. In an effort to determine which transactions are most prone to fraud, supervised fraud models were constructed. We cover the steps involved in exploring and cleaning data, creating variables, selecting features, methods for the model, and the outcomes. Logistic regression, neural networks, boosted trees, random forests, and support vector machines are the five supervised models that are investigated and evaluated. For this specific dataset, the boosted tree model demonstrates the most effective fraud detection result (FDR = 49.83%). Credit card fraud detection systems may make use of the final model. Insurance and telecommunications are two linked industries that may use a comparable model building method to identify or prevent fraud.

"A review and new perspectives on representation learning,"

Data representation is crucial to the performance of machine learning algorithms. We postulate that this is due to the fact that various representations may entangle and conceal many explanatory reasons of variation hidden within the data. While domain-specific information is useful for designing representations, generic previous learning is also an option, and the pursuit of artificial intelligence is driving the development of more effective representation-learning algorithms that exploit this kind of learning. Recent developments in deep learning and unsupervised feature learning are covered in this article. Topics covered include deep networks, probabilistic models, autoencoders, and manifold learning. This raises future open concerns about the best ways to learn representations, how to compute representations (inference), and the geometrical relationships among manifold learning, density estimation, and representation learning.

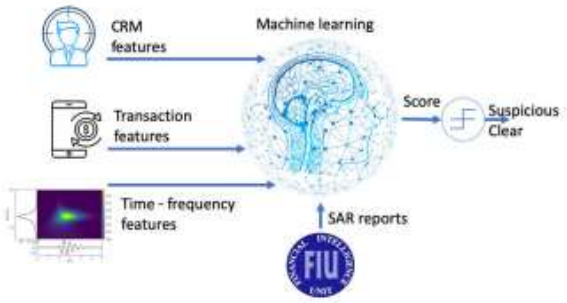
"What is feature engineering and how is it used in predictive modeling?"

Neural networks, decision trees, gradient boosting machines, and random forests are all examples of machine learning models that can take a feature vector as input and provide a prediction. Feature vectors with predicted outputs are supplied to these models so that they may learn in a supervised manner. Adding new features by hand using the existing feature set is standard procedure. The purpose of these manufactured features is to supplement or perhaps completely replace some of the current feature vector. The values of the other characteristics are used to compute these designed features, which are basically fields. The majority of the work in engineering such features is labor-intensive and done by hand. Furthermore, certain constructed characteristics will elicit distinct responses from various kinds of models. This article presents the results of empirical studies that show which kinds of

designed features work best with certain ML models. To do this, we create many datasets, each tailored to make use of a certain engineering characteristic. How well the machine learning model can independently synthesize the required characteristic is shown by the experiment. The provision of an engineering feature is superfluous if the model can synthesise it. According to the results, the examined models really do exhibit distinct behavior when endowed with different kinds of designed characteristics.

3. SYSTEM ANALYSIS

SYSTEM ARCHITECTURE



3.1 Existing System

Using just time-frequency features results in a false positive rate of 14.9 percent and an F-score of 59.05%. When transaction and CRM characteristics are used, the false positive rate decreases to 11.85% and the F-score increases to 74.06%.

Disadvantages of Existing System:

1. Less Prediction.
2. Security is less.

3.2 Proposed System

In order to repurpose illicitly acquired funds for legal transactions, a practice known as "money laundering" (ML) is used. A vast array of crimes (often referred to as predicate offences) are interconnected via ML. These include drug and human trafficking, terrorism, extortion, kidnapping for ransom, bribery, embezzlement, tax fraud, corruption, and many more. The International Monetary Fund (IMF) believes that financial institutions worldwide launder up to $2 trillion USD yearly, making ML one of the world's biggest markets. However, it is difficult to provide a precise evaluation of the scope of such a complex subterranean sector.

Advantages of Proposed System:

1. Security is more.

2. More Prediction.

4. Modules Information:

We have developed the following modules to carry out this project.

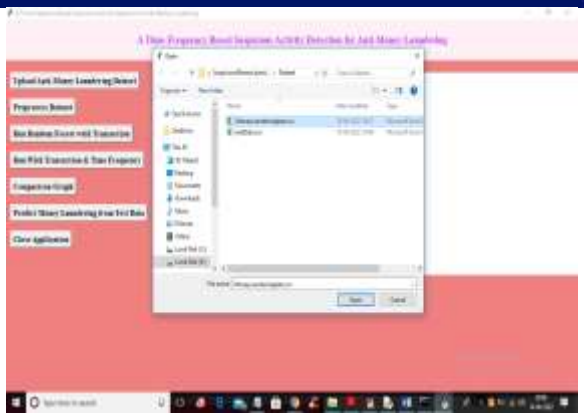
1) First, we'll use this module to upload an anti-money-laundering dataset. Then, we'll use the application to produce a graph that shows the number of regular and money-laundering transactions.
2) Preprocess Dataset: Since Random Forest only accepts numeric input, we must transform the dataset's missing and non-numeric elements into numeric ones using the Label Encoder class. No numeric data is ever lost while using a label encoder.
3) Calculate FSCORE, True Positive, and Negative Rate using Random Forest with Transaction: The processed data will be divided into train and test datasets. The program will utilize 80% of the dataset to train Random Forest on the transaction data. Then, it will apply this trained model to 20% of the dataset to determine these metrics.

4) Run with Transaction and Time Frequency: this module will be used to use the Fast Fourier Transform technique to transform transaction data into Time Frequency. Then, Random Forest will be used to train on this Time Frequency data in order to get FSCORE, TPR, and FPR values.
5) Comparison Graph: This module will be used to plot an FSCORE Random Forest graph comparing the two sets of data: time frequency FFT and transaction data.
6) Use the Random Forest model to determine whether the test data has typical transaction characteristics or money-laundering features. This module allows us to submit test data and use it to make predictions.

5. SCREEN SHOTS

Double click the "run.bat" file to launch the project and see the screen below.

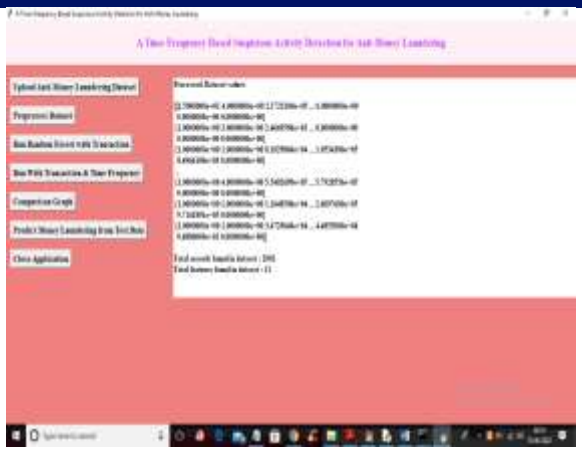
Click the "Upload Anti-Money Laundering Dataset" button on the top screen to upload the dataset and get the results shown below.



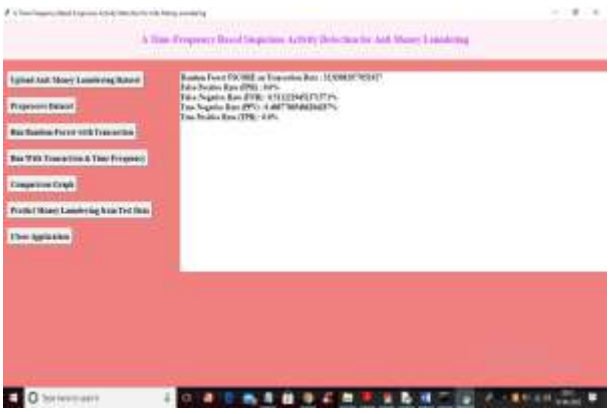
Choose the "Money Laundering" dataset in the above page, upload it, and then click the "Open" button to see the result below.

The loaded dataset is shown in the above screen. It comprises both numerical and non-numeric data, so we must preprocess the data before turning it into numerical data. In the graph above, the labels are represented by the x-axis, which indicates counts, and the labels are represented by the y-axis, which indicates money-laundering transactions (0 and 1), respectively. Close the above graph now, and then choose "Preprocess Dataset" to clean the dataset and get the result below.

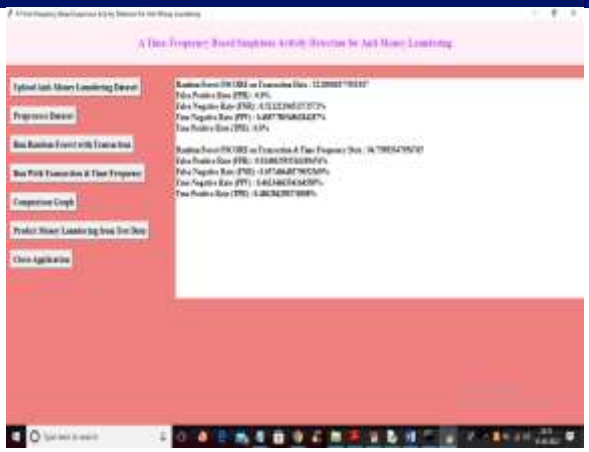
International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org



The dataset is ready when we click the "Run Random Forest with Transaction" button to train Random Forest on the transaction dataset and obtain the output shown below. The entire dataset is displayed in the above screen converted to numeric format. The last two lines of the dataset show that it contains 2001 records, each of which contains 11 features.

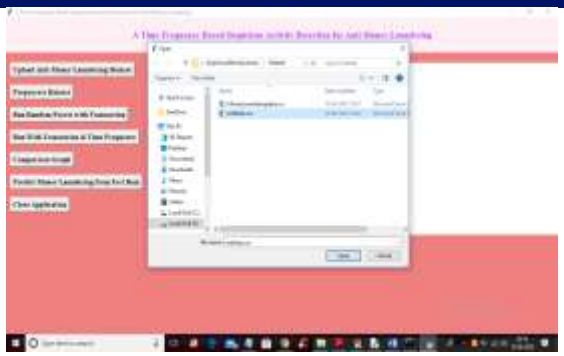


The accompanying screen shows that the True Positive Rate (TPR) was 0 and the FSCORE was 32% based on transaction data using Random Forest. The characteristics of the transaction data will thus prevent Random Forest from training properly. To rectify this, select the "Run With Transaction & Time Frequency" button, convert the transaction data to time frequency, and then retrain Random Forest to get the result shown below.



By transforming transaction data into time frequency, we can create an accurate prediction model. In the screen above, Random Forest using Time Frequency Transaction data obtained 94% FSCORE and TPR as 45%. To see the graph below, click the "Comparison Graph" button now.

The method name is shown on the x-axis in the above graph, and the FSCORE is shown on the y-axis. Time frequency increased for both techniques. Close the above graph, then click the "Predict Money Laundering from Test Data" button to submit test data and get the prediction outcome.



To get the output below, choose the "testData.csv" file in the top screen, upload it, and then click the "Open" button.

In above screen in square bracket we can see transaction test data and after arrow symbol =➔we got prediction output as 'Normal or Money Laundering'

6. CONCLUSION:

We have shown in this research that the inclusion of time-frequency features enhances the quality of the data science model and streamlines the feature selection procedure. For the first time, time-frequency characteristics including skewness, variance, mean, and Kurtosis have been used to machine learning model training for the identification of fraudulent transactions. Consequently, by figuring out the suggested time-frequency feature set, the feature engineering phase may be reduced. For the financial institutions, this might possibly save several person-months of modeling research. The great degree of correctness of the suggested method has been shown using actual financial data, and it has been implemented in Python. It is simple to modify the generic method to identify questionable transactions in other kinds of organizations. Time-frequency characteristics may differentiate between suspect and clear situations, increasing the transaction monitoring system's efficiency and AUC, according to an investigation of real customer data. Kurtosis offered the highest degree of differentiation in the model among the various time-frequency characteristics. Financial firms may save recurring penalties and human resources expenses of millions of dollars thanks to improvements in accuracy and the capacity to identify money laundering incidents that were previously undetectable. For frequency domain analysis, only a low complexity Fourier transform-based method is used in this study. The time-frequency analysis may be completed using additional linear and non-linear transformations as a future project. Comparing different window durations, increment sizes, and doing the research across other banking channels (such as ATMs, branches, and the web) might potentially provide benefits. Furthermore, the same research may be expanded to look at network properties as opposed to individual entities. Specifically, when a client has many accounts at various banks, only the FIUs are able to examine the whole picture. Therefore, it would also be advantageous to replicate this analysis with more FIU data. Time-frequency characteristics thus offer a wide range of possible applications in the field of financial behavior analysis in the future.

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.