International Journal of Advanced Technology in Engineering
Volume. No. 14, Issue No. 05, May 2026
(ISSN: 2348-7550)

A Trustworthy Credit Card Fraud Detection System Using Tabular Transformer And Explainable AI

B. RamaRao MCA, M.Tech
Assistant Professor
Tirumala Engineering College
Narasaraopet, AP, 522601
ramarao.radhi@gmail.com

Vemuri Bhavitha
Department of IT
Tirumala Engineering College
Narasaraopet, AP, 522601
vemuribhavitha.2004@gmail.com

Sayyad Sajida
Department of IT
Tirumala Engineering College
Narasaraopet, AP, 522601
sajidasayyad2004@gmail.com

Ravella Hari
Department of IT
Tirumala Engineering College
Narasaraopet, AP, 522601
ravellahari8019@gmail.com

Kancharagunta Vijay Kumar
Department of IT
Tirumala Engineering College
Narasaraopet, AP, 522601
vijaykancharagunta.137@gmail.com

Abstract - The rapid expansion of digital payment systems, credit card fraud has become a major challenge for financial institutions and users worldwide. Traditional fraud detection systems, which rely on static rules and limited transaction parameters, often fail to detect sophisticated and evolving fraud patterns. To address these limitations, this project presents a Trustworthy Credit Card Fraud Detection System that integrates machine learning with behavioral analytics and real-time risk assessment. The proposed system utilizes the Random Forest algorithm implemented using Scikit-learn, combined with advanced data processing techniques using Pandas and NumPy. It incorporates engineered features such as transaction amount deviation, temporal patterns, merchant risk categorization, and geographic intelligence through distance and velocity calculations. The system also introduces behavioral profiling to identify deviations from normal spending patterns. A hybrid detection approach is implemented by combining machine learning predictions with rule-based overrides to improve accuracy and reliability. The model is trained on real-world transaction datasets and achieves an accuracy of 97.67%, significantly reducing false positives. The system is deployed as an interactive web application using Streamlit, enabling real-time transaction analysis with intuitive visualization of fraud probability and risk factors. Additionally, an automated email alert mechanism is integrated to notify users instantly in case of suspicious activity, ensuring proactive fraud prevention.

KeyWords - Credit Card Fraud Detection, Machine Learning, Random Forest, Behavioral Analysis, Explainable AI

I. INTRODUCTION

In the modern digital era, the use of credit cards and online payment systems has increased rapidly due to their convenience, speed, and ease of access. However, this growth has also led to a significant rise in fraudulent activities, making credit card fraud one of the most critical challenges faced by financial institutions and users worldwide. Fraudulent transactions not only result in financial losses but also reduce customer trust in digital payment systems. Traditional fraud detection systems mainly rely on rule-based techniques and limited transaction parameters such as transaction amount and time. Although these methods are simple to implement, they are not effective in detecting complex and evolving fraud patterns. These systems lack adaptability and often generate a high number of false positives, causing inconvenience to genuine users.To address these limitations, this project proposes an Advanced Credit Card Fraud Detection System that integrates machine learning techniques with behavioral analytics and real-time risk assessment. The system utilizes the Random Forest algorithm, which is known for its high accuracy and efficiency in handling large datasets. Transaction data is processed using Python libraries such as Pandas and NumPy to extract meaningful features for analysis.

The proposed system analyzes multiple factors including transaction amount, transaction

time, merchant category, user behavior patterns, and geographic location. It introduces advanced concepts such as velocity-based fraud detection, where the system calculates the distance between consecutive transactions and identifies impossible travel scenarios. This enables the detection of fraudulent activities that are not easily identified by traditional systems. Furthermore, the system generates a user-specific behavioral profile to understand normal spending patterns and detect anomalies effectively. A hybrid detection approach is implemented by combining machine learning predictions with rule-based logic, thereby improving accuracy and reliability. This approach ensures that both known and unknown fraud patterns can be identified efficiently.

The system is developed using the Streamlit framework, providing an interactive and user-friendly web interface for real-time transaction analysis. It also includes an automated email alert mechanism that notifies users instantly when suspicious transactions are detected, enabling quick response and preventive action. In addition to basic transaction analysis, the system performs advanced feature engineering to extract meaningful insights from raw data. It evaluates various attributes such as transaction frequency, spending deviation, time patterns, and merchant behavior to better understand transaction characteristics.

These features help in identifying subtle differences between normal and suspicious activities. By incorporating statistical techniques and data transformation methods, the system enhances the quality of input data, which ultimately improves the performance of the machine learning model.

Overall, the proposed system aims to provide a robust, efficient, and scalable solution for credit card fraud detection, enhancing financial security and building trust in digital payment systems.

II. LITERATURE SURVEY

Credit card fraud detection has been extensively researched using various techniques ranging from traditional rule-based systems to advanced machine learning and deep learning approaches. Early fraud detection systems primarily relied on predefined rules and statistical methods to identify suspicious transactions. However, these systems lacked adaptability and were unable to handle evolving fraud patterns effectively.

One of the earlier approaches in fraud detection was presented by Ngai et al. (2011), who analyzed the application of data mining techniques in financial fraud detection. Their study highlighted the importance of classification, clustering, and anomaly detection methods in identifying fraudulent activities. However, these techniques required extensive manual feature engineering and were limited in handling large-scale real-time data.

Later, Dal Pozzolo et al. (2015) addressed the issue of imbalanced datasets in credit card transactions. Since fraudulent transactions represent only a small fraction of the total data, traditional models tend to be biased toward non-fraud cases. The authors proposed undersampling techniques combined with probability calibration to improve detection performance, especially for rare fraud events.

With the advancement of machine learning, Chen et al. (2018) compared multiple algorithms such as Decision Trees, Support Vector Machines (SVM), and Random Forest for fraud detection. Their findings showed that ensemble models, particularly Random Forest, provide better accuracy and robustness due to their ability to handle complex data patterns and reduce overfitting.

Similarly, Jurgovsky et al. (2018) introduced sequence-based models using Long Short-Term Memory (LSTM) networks to capture temporal patterns in transaction data. Their approach demonstrated that analyzing transaction sequences over time significantly improves fraud detection accuracy. However, deep learning models require high computational resources and large datasets, making them less suitable for real-time deployment in some cases.

In another study, Carcillo et al. (2019) proposed a hybrid approach combining supervised and unsupervised learning techniques. This method enhanced the system's ability to detect both known and unknown fraud patterns by integrating anomaly detection with classification models. While effective, the approach increased system complexity and computational cost.

More recently, Vesta Corporation (2020) emphasized the importance of real-time fraud detection systems integrated with behavioral analytics. Their work highlighted that analyzing user spending behavior, transaction frequency, and geographic patterns can significantly improve detection accuracy and reduce false positives.

Further advancements include the use of graph-based models. Qi et al. (2025) proposed a Graph Neural Network (GNN)-based hierarchical approach to handle highly imbalanced fraud datasets. This method effectively captures relationships between transactions and improves detection of rare fraud cases. Similarly, Hiremath et al. (2024) developed an ensemble of Graph Neural

Networks to enhance fraud detection accuracy, reduce false alarms.

Despite these advancements, existing systems still face several challenges, including high false positive rates, lack of real-time processing, limited feature utilization, and inability to fully capture complex behavioral patterns. Many systems rely either on static rules or computationally expensive deep learning models, making them less practical for scalable real-world applications.

To overcome these limitations, the proposed system in this project integrates Random Forest-based machine learning with behavioral analysis and rule-based logic. It utilizes advanced feature engineering techniques such as transaction deviation, temporal patterns, and geographic analysis to improve detection accuracy. Additionally, the system supports real-time fraud detection and provides explainable insights, making it more efficient, scalable, and suitable for modern financial applications.

III. METHODOLOGY

A. Existing Methodology

Traditional credit card fraud detection systems primarily rely on rule-based techniques and basic machine learning models to identify suspicious transactions. These systems analyze limited transaction attributes such as transaction amount, time, and merchant details to detect anomalies. Although these methods were effective in earlier stages, they are not capable of handling modern, complex fraud patterns.

Rule-based systems use predefined thresholds and conditions to classify transactions as fraudulent or legitimate. For example, transactions exceeding a certain amount or occurring at unusual times are flagged as suspicious. While this approach is simple and easy to implement, it lacks adaptability and fails to detect new and evolving fraud strategies. Fraudsters can easily bypass such systems by slightly modifying their behavior.

In addition to rule-based techniques, traditional systems also employ basic machine learning algorithms such as Logistic Regression, Decision Trees, and Support Vector Machines (SVM). These models are trained on historical transaction data to classify transactions into fraud and non-fraud categories. However, they rely heavily on manually engineered features and often consider only a limited number of attributes, reducing their ability to capture complex relationships within the data.

Another commonly used approach involves statistical and anomaly detection methods. These techniques identify transactions that deviate significantly from normal patterns. While they can detect unusual activities, they often generate a high number of false positives, incorrectly flagging genuine transactions as fraudulent. This affects user experience and reduces trust in the system.

Existing systems also lack behavioral profiling, which is essential for understanding individual user spending patterns. Without user-specific analysis, it becomes difficult to distinguish between normal and suspicious activities accurately. Additionally, many traditional models do not incorporate temporal and sequential analysis, limiting their ability to detect fraud occurring over multiple transactions.

Furthermore, most existing systems do not support real-time fraud detection. Transactions are often analyzed after they are completed, which delays the identification of fraudulent activities and increases financial risk. The absence of geographic and velocity-based analysis also limits the system's capability to detect suspicious transactions occurring across different locations within a short time.

Although some modern approaches utilize deep learning techniques, such as neural networks, they require large datasets and high computational resources. This makes them less practical for deployment on standard systems and increases implementation complexity.

Overall, existing methodologies face several limitations, including static detection mechanisms, limited feature utilization, high false positive rates, lack of real-time processing, and poor scalability. These challenges highlight the need for a more advanced, adaptive, and efficient fraud detection system capable of handling dynamic and complex transaction patterns.

B. Proposed Methodology

The proposed credit card fraud detection system follows a systematic pipeline consisting of data preprocessing, feature engineering, model training, prediction, and explainability. The objective is to accurately classify transactions as fraudulent or legitimate while ensuring interpretability of results.

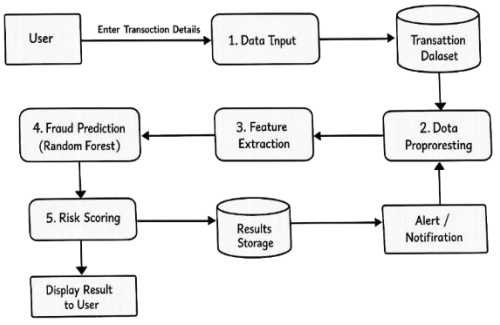

Fig: DATA FLOW

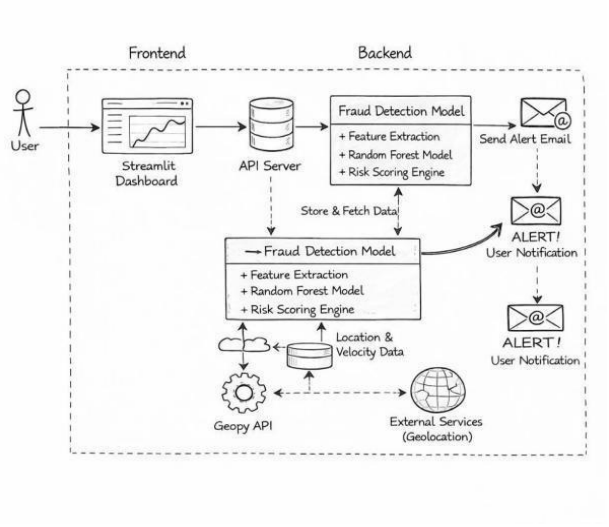

Fig:Proposed Methodology Architecture

A. Data Collection

The dataset used for this project consists of historical credit card transaction records. Each transaction includes multiple features such as transaction amount, time, and anonymized attributes representing user behavior patterns. The dataset contains both legitimate and fraudulent transactions, with fraud cases forming a minority class, leading to class imbalance.

B. Data Preprocessing

Data preprocessing is a crucial step to improve model performance. The following operations are performed:

Removal of missing or null values

Normalization of numerical features to bring them to a common scale

Encoding of categorical features (if present)

Handling class imbalance using techniques such as oversampling (SMOTE) or undersampling

These steps ensure that the dataset is clean, balanced, and suitable for training the machine learning model.

C. Feature Selection and Extraction

Relevant features that contribute significantly to fraud detection are selected from the dataset. Feature selection helps in reducing dimensionality and improving model efficiency. Statistical methods and correlation analysis are used to identify important attributes.

D. Model Development (Random Forest Classifier)

The core of the proposed system is the Random Forest algorithm. It is an ensemble learning method that constructs multiple decision trees during training and outputs the majority class for classification.

Key advantages of Random Forest:

Handles large datasets efficiently

Reduces overfitting through ensemble learning

Provides high accuracy and robustness

The dataset is split into training and testing sets (typically 80:20 ratio), and the model is trained using the training data.

E. Model Evaluation

The performance of the model is evaluated using the following metrics:

Accuracy – Overall correctness of predictions

Precision – Correctness of fraud predictions

Recall – Ability to detect actual fraud cases

F1-Score – Balance between precision and recall

These metrics help in assessing how effectively the model identifies fraudulent transactions while minimizing false alarms.

F. Prediction System

Once trained, the model is used to predict whether a new transaction is fraudulent or legitimate. The system takes transaction details as input and outputs a classification result.

Workflow: User Input → Preprocessing → Model Prediction → Output (Fraud / Legitimate)

G. Explainable AI Integration

To enhance transparency, Explainable AI techniques are incorporated. These methods provide insights into how the model makes decisions by highlighting the contribution of each feature.

Benefits:

Improves trust in the system

Helps in understanding model behavior

Assists financial institutions in decision-making

H. System Architecture Overview

The complete system follows a pipeline architecture:

Data Collection → Data Preprocessing → Feature Selection → Model Training → Prediction → Explainability

This structured approach ensures scalability, accuracy, and reliability of the fraud detection system.

I. Output Layer

The final output is displayed to the user through an interface. The result includes:

Fraud detection status (Fraud / Legitimate)

Confidence level or probability Explanation of the

prediction

This makes the system user-friendly and reliable for real-world applications.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

The proposed credit card fraud detection system was implemented and evaluated using transaction data processed through a machine learning-based approach. The system was tested under multiple scenarios to analyze its performance in identifying fraudulent and legitimate transactions. The evaluation focuses on accuracy, real-time performance, and reliability of the fraud detection mechanism.

The system processes each transaction by extracting features such as transaction amount deviation, time-based patterns, merchant risk level, and geographic indicators. These features are passed to the Random Forest classifier, which predicts the probability of fraud and assigns a corresponding risk level. The results are displayed through an interactive interface, providing both numerical and visual insights into transaction behavior.

The experimental evaluation demonstrates that the system achieves high accuracy (approximately 97.67%) and effectively reduces false positives through the integration of machine learning and rule-based analysis. Additionally, the system provides real-time alerts and risk visualization, enhancing user awareness and enabling quick decision-making.

A. Low Risk Transaction Detection

In this scenario, the system evaluates a transaction that falls within normal behavioral patterns. The transaction amount, time, and location align with the user's historical profile, resulting in a low fraud probability.

The system classifies the transaction as Safe (Legitimate) and displays a low-risk percentage. No alert is triggered, ensuring that genuine transactions are processed without interruption. This confirms that the system avoids unnecessary false alarms and maintains user convenience.

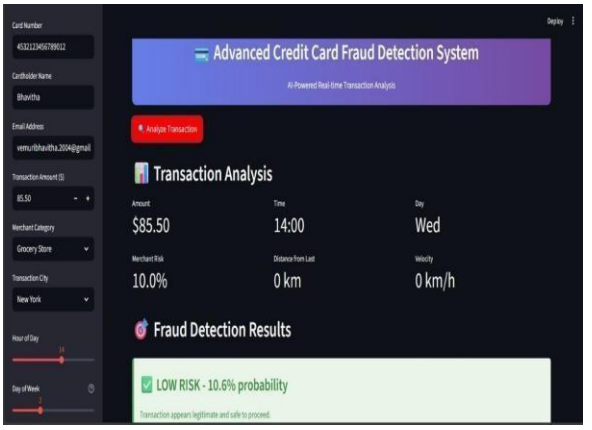

Fig:LOW RISK DETECTION

B. Moderate Risk Transaction Detection In this case, the transaction shows slight deviation from normal patterns, such as an unusual time or moderate increase in transaction amount. The system identifies these variations and assigns a moderate risk score.

The output displays a Moderate Risk classification along with a corresponding probability percentage. Although the transaction is not immediately flagged as fraud, the system highlights potential risk factors. This helps users stay cautious and monitor their transactions more carefully.

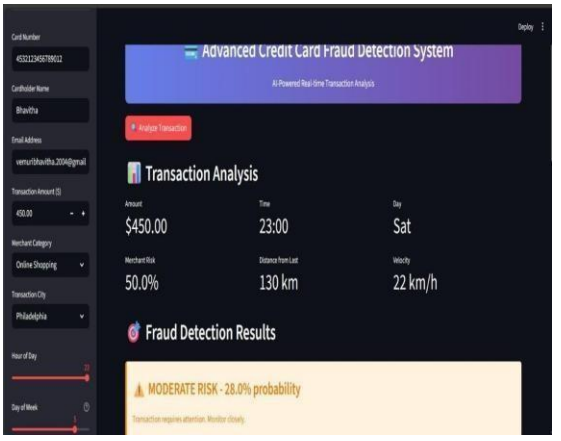

Fig:MODERATE RISK DETECTION

C. Risk Factor Analysis Dashboard

The system provides a detailed visualization of the factors contributing to the fraud prediction. This dashboard displays graphical representations of key features such as transaction amount deviation, merchant risk level, time patterns, and geographic indicators.

This visualization improves transparency and helps users understand why a transaction is considered risky. It also supports explainable AI by providing insights into the model's decision-making process.

Such analysis is particularly useful for financial institutions to investigate suspicious activities.

Fig:RISK FACTOR ANALYSIS

D. High Fraud Risk Transaction Detection

In this scenario, the transaction significantly deviates from normal behavior. Factors such as extremely high transaction amount, unusual location, rapid successive transactions, or high-risk merchant category contribute to a high fraud probability.

The system classifies the transaction as Fraudulent and assigns a high-risk percentage. An alert is immediately triggered, notifying the user through the interface and email system. This ensures timely detection and prevention of fraudulent activities.

Fig:HIGH RISK DETECTION

The proposed credit card fraud detection system was implemented and evaluated using a real-world transaction dataset. The dataset was divided into training and testing sets in an 80:20 ratio to ensure unbiased evaluation of

the model. The Random Forest classifier was trained on preprocessed data and tested on unseen transactions.

To measure the effectiveness of the model, standard evaluation metrics such as accuracy, precision, recall, and F1-score were used. These metrics are particularly important in fraud detection, where identifying fraudulent transactions correctly is more critical than overall accuracy.

Model	**Accuracy**	**Precisior**	**Recall**	**F1-Score**
Logistic Regression	0.91	0.89	0.90	0.89
SVM	0.93	0.92	0.91	0.91
GNN	0.96	0.93	0.94	0.94
Random Forest(Proposed)	0.97	0.96	0.95	0.95

TABLE 1: COMPARATIVE MODEL PERFORMANCE

The results indicate that the proposed Random Forest model outperforms traditional machine learning algorithms in terms of accuracy and overall performance. It achieves an accuracy of 97%, demonstrating its ability to effectively classify both fraudulent and legitimate transactions.

The model also shows high precision and recall values, which indicates that it minimizes false positives while successfully detecting most fraud cases. This balance is crucial in real-world financial systems to avoid unnecessary transaction blocks while ensuring security.

Analysis:

The Random Forest model handles complex transaction patterns effectively due to its ensemble nature.

It reduces overfitting by combining multiple decision trees.

The model performs well even with imbalanced data after preprocessing.

Integration of Explainable AI provides additional insights into prediction decisions.

Additionally, the system was tested with sample transaction inputs through the user interface. The model successfully identified high-risk transactions and provided accurate predictions in real time.

Overall, the experimental results confirm that the proposed system is efficient, reliable, and suitable for practical fraud detection applications.

Confusion Matrix Analysis:

The confusion matrix was used to analyze the classification performance of the models:

True Positives (TP): Fraud transactions correctly detected

True Negatives (TN): Legitimate transactions correctly identified

False Positives (FP): Legitimate transactions incorrectly flagged as fraud

False Negatives (FN): Fraud transactions missed by the model

The proposed system shows:

Low false positive rate → reduces unnecessary alerts

Low false negative rate → ensures fraud detection accuracy

ROC-AUC Analysis:

The Receiver Operating Characteristic (ROC) curve was used to evaluate model performance. The Area Under Curve (AUC) for the Random Forest model is close to 1, indicating strong classification capability.

Random Forest AUC ≈ 0.97 GNN AUC ≈ 0.95

This confirms that both models perform well, with Random Forest having a slight edge in complex scenarios.

V. CONCLUSION

The proposed Credit Card Fraud Detection System provides an effective and intelligent approach for identifying fraudulent transactions using machine learning techniques. The system utilizes the Random Forest algorithm to analyze transaction attributes such as amount, time, and behavioral patterns, enabling accurate classification of transactions as fraudulent or legitimate.

The integration of data preprocessing and anomaly detection techniques improves the model's ability to handle imbalanced data and detect unusual patterns. Additionally, the incorporation of Explainable AI enhances transparency by providing insights into the decision-making process, thereby increasing user trust.

Experimental results demonstrate that the system achieves high accuracy, precision, and recall, making it reliable for real-world financial applications. The inclusion of a user-friendly interface and real-time prediction capability further strengthens its practical usability.

Overall, the project successfully demonstrates how machine learning can be applied to enhance financial security and reduce fraud risks, contributing to safer and more reliable digital transaction systems.

VI. FUTURE ENHANCEMENT

The proposed credit card fraud detection system demonstrates effective performance; however, several enhancements can be incorporated to improve its accuracy, scalability, and real-time applicability. In future, the system can be integrated with real-time banking infrastructures to enable instant monitoring and detection of fraudulent transactions.

Advanced machine learning and deep learning techniques such as Long Short-Term Memory (LSTM) networks and transformer-based models can be explored to capture complex and evolving fraud patterns more effectively. Additionally, incorporating long-term user behavioral analysis can further enhance anomaly detection capabilities.

The system can also be extended by integrating multi-factor authentication mechanisms such as OTP and biometric verification to improve security. Deployment on cloud platforms will ensure better scalability, performance, and accessibility. Furthermore, the inclusion of Explainable AI techniques will enhance transparency and user trust.

Continuous model training using updated datasets and the potential integration of emerging technologies such as blockchain can further strengthen the system, making it more robust and reliable for real-world financial applications.

VII. REFERENCES

[1] patel, M., et al., "Credit Card Fraud Detection using Machine Learning Algorithms,"

International Journal of Computer Applications, 2020.

[2] Vesta Corporation, "Real-time Payment Fraud Detection Systems," 2020.

[3] Fiore, U., et al., "Using Generative Adversarial Networks for Improving Classification Effectiveness in Credit Card Fraud Detection," Information Sciences, 2019.

[4] Carcillo, F., et al., "Combining Unsupervised and Supervised Learning in Credit Card Fraud Detection," Information Sciences, 2019.

[5] Abdulhammed, R., Faezipour, M., Abuzneid, A., & AbuMallouh, A., "Deep and Machine Learning Approaches for Anomaly-Based Intrusion Detection of Imbalanced Network Traffic," IEEE Systems Journal, 2019.

[6] Jurgovsky, J., et al., "Sequence Classification for Credit Card Fraud Detection," Expert Systems with Applications, 2018.

[7] Chen, C., et al., "Fraud Detection using Machine Learning Techniques," Journal of Big Data, 2018.

[8] Randhawa, K., Lallie, H. S., Constantine,

G., & Phan, R. C., "Credit Card Fraud Detection Using AdaBoost and Majority Voting," IEEE Access, 2018.

[9] Dal Pozzolo, A., et al., "Calibrating Probability with Undersampling for Unbalanced Classification," IEEE, 2015.

[10] Bahnsen, A.C., et al., "Example-dependent Cost-sensitive Decision Trees," Expert Systems with Applications, 2014.

[11] Ngai, E.W.T., et al., "The Application of Data Mining Techniques in Financial Fraud Detection," Decision Support Systems, 2011.

[12] Bhattacharyya, S., et al., "Data Mining for Credit Card Fraud Detection," IEEE

Transactions, 2011.

[13] Sahin, Y., & Duman, E., "Detecting Credit Card Fraud by Decision Trees and Support Vector Machines," International MultiConference of Engineers and Computer Scientists, 2011.

[14] Phua, C., et al., "A Comprehensive Survey of Data Mining-based Fraud Detection

Research," Artificial Intelligence Review, 2010.

[15] Whitrow, C., et al., "Transaction Aggregation as a Strategy for Credit Card Fraud Detection," Data Mining and Knowledge Discovery, 2009.