

**BINDURA UNIVERSITY OF SCIENCE EDUCATION**  
**FACULTY OF SCIENCE AND ENGINEERING**



**Department of Mathematics and Statistics**

**Detection And Prevention Of Financial Fraud In Banking Transactions Using Neural  
Networks: A Case Of Cabs Bank**

**BY**

**Melissa Marimbi**

**B190599B**

***A PROJECT SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS OF  
BACHELOR OF SCIENCE HONORS DEGREE IN STATISTICS AND FINANCIAL  
MATHEMATICS***


**SUPERVISOR**

**Mr B. Kusotera**

**June 2024**

### **DECLARATION OF AUTHORSHIP**

I MELISSA MARIMBI declare that this research project is my original work and has not been copied or extracted from previous sources without due acknowledgement of the sources


Signature .....

Date...10/06/2024.....

### APPROVAL FORM

I, Melissa Marimbi do hereby declare that this submission is my own work apart from the references of other people's work, which has duly been acknowledged. I hereby declare that this work has not been presented neither in whole nor in part for any degree at this university or elsewhere.

Melissa Marimbi




10/06/2024

Signature

Date

Certified by:

Mr. B. Kusotera (Supervisor)



10/06/2024

Signature

Date

Dr. M. Magodora (Chairperson)



20/06/2024

Signature

Date

## **DEDICATION**

This study is dedicated to my family whose unwavering support and encouragement have been my bedrock throughout this journey.

To my parents, who instilled in me the value of education and perseverance. Your love and guidance have been the foundation of my achievements.

Thank you for believing in me and inspiring me to strive for excellence.

## **ACKNOWLEDGEMENTS**

I would like to express my deepest gratitude to all those who have supported and guided me throughout the completion of this study.

First and foremost, I wish to thank my supervisor, [Mr Kusotera], for his invaluable guidance, insightful feedback, and unwavering support. Your expertise and encouragement have been crucial in shaping this research.

I am immensely grateful to the management and staff of CABS Bank Zimbabwe for their cooperation and for providing the necessary data and resources. Your willingness to participate and share insights made this study possible.

Special thanks to my colleagues and peers, whose constructive criticism and suggestions have significantly enriched this work. Your collaborative spirit and intellectual contributions have been greatly appreciated.

I also extend my heartfelt thanks to my family for their unwavering support, patience, and encouragement throughout this journey. Your love and understanding have been my constant source of motivation.

Finally, I would like to acknowledge the contributions of all those whose names may not appear here but have supported me in various ways. Your kindness and assistance have been invaluable.

Thank you all for your unwavering support and belief in my potential.

## ABSTRACT

Financial fraud poses a significant threat to the stability and integrity of banking institutions worldwide. With the increasing volume and complexity of digital transactions, traditional rule-based fraud detection systems are often inadequate in identifying sophisticated fraud patterns. This study explores the application of neural networks to detect and prevent financial fraud in banking transactions at CABS Bank Zimbabwe. The research employs a mixed-methods approach, combining quantitative analysis of transaction data with qualitative insights from bank employees and customers. Transactional data from CABS Bank is used to train and evaluate a neural network model, focusing on features such as transaction amount, frequency, geographical location, and timing. The model's performance is compared with existing fraud detection methods, demonstrating significantly higher accuracy and reliability in detecting fraudulent transactions. Key findings reveal that neural networks can effectively identify complex fraud patterns and process transactions in real-time, providing immediate alerts for suspicious activities. The study also highlights the importance of feature engineering and continuous model updates to adapt to evolving fraud tactics. Interviews with bank employees underscore the need for integrating advanced fraud detection systems into existing workflows and ensuring staff training for optimal utilization. The implementation of neural network-based fraud detection at CABS Bank promises to enhance the bank's security measures, reduce financial losses, and improve customer trust. This study concludes with practical recommendations for the bank, including adopting neural networks, enhancing data collection, and conducting regular employee training. Suggestions for future research include exploring advanced models, improving model interpretability, and examining the scalability of neural network systems. By leveraging the power of neural networks, CABS Bank Zimbabwe can significantly bolster its fraud detection capabilities, safeguarding financial assets and reinforcing customer confidence.

**Keyword:** *Financial Fraud, Neural Networks, Fraud Detection, Banking Transactions, Machine Learning, CABS Bank Zimbabwe, Data Preprocessing*

## TABLE OF CONTENTS

DECLARATION OF AUTHORSHIP .....	i
APPROVAL FORM .....	ii
DEDICATION .....	iii
ACKNOWLEDGEMENTS .....	iv
ABSTRACT .....	v
LIST OF FIGURES .....	ix
LIST OF TABLES .....	x
LIST OF ABBREVIATIONS .....	xi
CHAPTER 1: .....	1
INTRODUCTION OF STUDY .....	1
1.1 Introduction .....	1
1.2 Organisation of study .....	1
1.3 Background of the study .....	1
1.4 Overview of Central African Building Society (CABS) Bank .....	2
1.5 Statement of Problem .....	2
1.6 Research Objectives .....	3
1.7 Research Questions .....	3
1.8. Scope of the Study .....	4
Geographical Scope: .....	4
1.9 Limitations: .....	4
1.10 Significance of the Study .....	4
1.11 Assumptions of the Study .....	6
1.13 Definition of Terms .....	6
1.14 Chapter Summary .....	7
CHAPTER 2: .....	8
LITERATURE REVIEW .....	8
2.1 Introduction .....	8
2.2 Theoretical literature .....	8
2.2.1 Fraud detection .....	8
2.2.2 Types of Financial Fraud in banking transactions .....	8
2.2.3 Traditional Fraud Detection Techniques .....	10
2.2.4 Limitations .....	11

2.2.5 Emergence of Neural Networks .....	11
2.2.6 Neural networks in fraud detection .....	13
2.2.7 Population size .....	13
2.2.8 Theoretical review of fraud detection techniques and algorithms.....	14
2.2.9 Neural network classifier .....	14
2.2.10 Types of neural network classifiers .....	14
2.3 Empirical literature.....	16
2.3.1 Related studies.....	16
2.4 Research Gap .....	17
2.5 Synthesis of Research Gaps.....	18
2.6 Proposed Conceptual Model.....	18
2.7 Advantages of neural networks .....	20
2.8 Limitations.....	20
2.9 Chapter summary.....	21
CHAPTER 3: .....	23
RESEARCH METHODOLOGY.....	23
3.1 Introduction .....	23
3.2 Research Design .....	23
3.2 Data Sources .....	23
3.3 Target Population and Sampling Procedures .....	24
3.4 Methods for Data Collection.....	24
3.5 Variables and Expected Relationships.....	24
3.6 Diagnostic Tests.....	24
3.6.1 Data quality checks .....	24
3.6.2 Feature correlation analysis.....	25
3.6.3 Multicollinearity analysis .....	25
3.7 Analytical Model .....	25
3.8 Neural network architecture.....	25
3.9 Model Validation (Fitness) Tests .....	27
3.10 Cross-Validation .....	27
3.11 Ethical Considerations.....	28
3.12 Chapter summary.....	28
CHAPTER 4: .....	29



DATA PRESENTATION, ANALYSIS AND INTERPRETATION .....	29
4.1 Introduction .....	29
4.2 Descriptive statistics.....	29
4.3 Pre-tests/Diagnostic tests .....	32
4.4 Test for multi-collinearity .....	32
4.5 Correlation of Variables.....	33
4.6 Model presentation .....	34
4.6.1 Model architecture .....	34
4.6.2 Training Process .....	34
4.7 Model output/Results .....	35
4.8 Prediction of Fraudulent Transactions .....	38
4.8.1 Transaction number 1 index 134453.....	39
4.8.2 Transaction number 2 index 135149.....	39
4.8.3 Transaction number 3 index 97238.....	39
4.8.4 Transaction number 4 index 134536.....	40
4.8.5 Transaction number 5 index 131525.....	40
4.9 Model validation tests/Model fitness tests .....	40
4.9.1 K-fold cross-validation.....	40
4.10 Discussion of findings.....	41
4.11 Implications of Research Findings for Fraud Detection in Banking.....	42
4.12 Chapter summary.....	42
CHAPTER 5: .....	43
SUMMARY, CONCLUSIONS AND RECOMMENDATIONS .....	43
5.1 Introduction .....	43
5.2 Summary of Findings.....	43
5.3 Conclusions .....	44
5.4 Recommendations .....	45
5.5 Areas for Further Research .....	47
5.6 Chapter Summary .....	48
REFERENCES.....	49
APPENDICES .....	52

## **LIST OF FIGURES**

Figure 2.1 proposed conceptual model .....	21
Figure 3.1 neural network architecture .....	27
Figure 4.1 Histogram for Amount distribution .....	<b>Error! Bookmark not defined.</b>
Figure 4.2 Pie chart for Class distribution .....	<b>Error! Bookmark not defined.</b>
Figure 4.3 Bar graph for Transaction Type against Class .....	<b>Error! Bookmark not defined.</b>
Figure 4.4 Correlation Heatmap for Time, Amount and Class .....	33
Figure 4.5 model parameter setting .....	34
Figure 4.6 model parameter setting .....	34
Figure 4.7 model training.....	35
Figure 4.8 Receiver Operating Curve (ROC) .....	37
Figure 4.9 Confusion Matrix – Neural Network.....	38
Figure 4.10 Next 5 fraudulent transactions predictions .....	39

## LIST OF TABLES

Table 4.1 Descriptive statistics. ....	29
Table 4.2 Output of VIF for independent variables .....	33

Table 4.3 Metrics for Neural networks model in predicting Fraudulent and Legit transactions ..	35
Table 4.4 Cross validation results .....	41
Table 4.5 Metrics outputs for Neural Network against traditional models..	<b>Error! Bookmark not defined.</b>

## LIST OF ABBREVIATIONS

<b>AI</b>	Artificial Intelligence
-----------	-------------------------

<b>ANN</b>	Artificial Neural Network
<b>APT</b>	Advanced Persistent Threat
<b>ATM</b>	Automated Teller Machine
<b>AUC</b>	Area Under the Curve
<b>BEC</b>	Business Email Compromise
<b>CABS</b>	Central African Building Society
<b>CNN</b>	Convolutional Neural Network
<b>CNP</b>	Card Not Present
<b>FNN</b>	Feedforward Neural Network
<b>LSTM</b>	Long Short-Term Memory
<b>PIN</b>	Personal Identification Number
<b>POS</b>	Point of Sale
<b>RNN</b>	Recurrent Neural Network
<b>ROC</b>	Receiver Operating Characteristic
<b>SMOTE</b>	Synthetic Minority Over-sampling Technique
<b>VIF</b>	Variance Inflation Factor

## **CHAPTER 1:**

### **INTRODUCTION OF STUDY**

#### **1.1 Introduction**

These days, technology is driving a massive revolution in the banking sector by making financial transactions easier to access and more effective. However, this digital revolution has also escalated the threat of financial fraud. As financial transactions increasingly move to online platforms, traditional rule-based methods of fraud detection are proving inadequate, necessitating a fundamental shift in security measures.

The financial sector, integral to economic stability and growth, facilitates trade, investment, and economic activities. Yet, it is also highly susceptible to various forms of fraud, such as identity theft, transaction fraud, money laundering, and phishing. These fraudulent activities pose significant risks to both banks and their customers. Traditional fraud detection systems are often unable to keep up with the sophisticated and evolving techniques used by fraudsters, underscoring the need for more advanced and adaptive security solutions.

This study examines the application of neural networks to detect and prevent financial fraud in banking transactions at the Central African Building Society (CABS) bank in Zimbabwe. By leveraging cutting-edge machine learning techniques, particularly neural networks, CABS aims to significantly improve its fraud detection capabilities, thereby ensuring the security of transactions and maintaining customer trust.

#### **1.2 Organisation of study**

This study is divided into five chapters, each addressing different aspects of the research on using neural networks to detect and prevent financial fraud in banking transactions at CABS Bank Zimbabwe. Chapter 1 which is an introductory chapter provides a concise overview, encapsulating the background, statement of the problem, research objectives, research questions, scope, significance, assumptions, limitations, definition of terms, and a conclusion. Chapter 2 (literature review), delves into the relevant literature on neural networks and financial fraud detection and prevention, constructing a theoretical framework. Chapter 3 delineates the research methodology, encompassing research design, data collection procedure, and data analysis techniques. Chapter 4 shows results findings of the study, and Chapter 5 concludes by summarizing key findings,

discussing implications, and offering recommendations for future research and practical application

### **1.3 Background of the study**

The banking industry is a vital channel for delivering financial services to consumers, corporations, and governments, and it plays a major role in the global economy. Banks are essential financial entities that take deposits from their clients and use the money to provide credit, loans, and other financial services. The complex function that banks play in the economy includes facilitating financial transactions, promoting economic growth, and skillfully managing a variety of hazards. (Vaquero, 2023)

Financial fraud incidents in the banking industry have increased since the advent of internet, mobile, and fintech businesses. As technology evolves, so does the technology employed by fraudsters, thereby changing the methods they use to carry out fraudulent activities" (Shaikh & Karjaluoto, 2016; Shaikh & Glavee-Geo, 2019; Zhang, Xue, & Dhillon, 2020).

Financial fraud in banking transactions refers to illegal activities involving the manipulation or deceit of banking systems for financial gain. it can also be defined as "intentional deception to secure unfair or unlawful gain", (Yang, 2020). This encompasses a wide range of fraudulent behaviors, from simple theft to complex schemes involving multiple parties and sophisticated technologies.

Detectors and fraudulent transactions have historically worked together. Fraudulent transactions are the main source of financial losses, and they happen more often than ever, particularly in the contemporary Internet era. Transaction fraud cost the economy over \$28 billion in 2019, \$30 billion in 2020, and more than \$32 billion in 2021. It is estimated that by 2022, global transaction fraud will have increased annually to \$34 billion. (Alwadain, 2022)

Thus, banks and other financial service providers may need an automated fraud detection solution in order to identify and filter financial activities. The goal of fraud detection systems is to separate anomalous behavior patterns from a large volume of transaction records, and then utilize those patterns to identify or monitor new transactions. Fraud detection and prevention are ongoing, cyclical processes that include monitoring, identifying, making choices, managing cases, learning, and incorporating the lessons learned into the system. (Omollo, 2020)

Neural networks have become a viable substitute for card fraud detection in the past few years., (Khan, 2022))Neural networks are able to detect patterns in data that are difficult to detect with



traditional methods, and have shown great promise in detecting and preventing card fraud (Larson, 2020). For example, neural networks have been used to detect suspicious transactions in real time, and to identify patterns of card fraud that would be difficult to detect with rule-based systems. Neural networks have also been used to create risk profiles of cardholders, which can be used to flag potentially fraudulent transactions. However, there are challenges associated with using neural networks for card fraud detection, including the need for large amounts of training data and the risk of overfitting (Esenogho, 2022)

#### **1.4 Overview of Central African Building Society (CABS) Bank**

The Central Africa Building Society (CABS) plays a pivotal role in the financial sector, contributing significantly to the economic landscape of the Central African region. Established with a mission to provide comprehensive banking services, CABS has evolved into a cornerstone institution that facilitates economic growth and financial inclusion.

CABS serves as a multifaceted financial institution, offering banking products and services to people, companies, and government entities. Its role extends beyond traditional banking to encompass diverse financial activities, including savings and deposit accounts, loans, investment services, and electronic banking solutions.

As a key player in the financial sector, CABS contributes to the stability and development of the region's economy. Its strategic initiatives focus on fostering financial literacy, empowering entrepreneurs, and supporting economic ventures that drive progress. Through innovative financial products and a commitment to customer satisfaction, CABS has positioned itself as a trusted partner for individuals and businesses seeking reliable and forward-thinking banking solutions.

In an era marked by technological advancements and evolving customer expectations, CABS stands at the forefront of digital transformation within the financial sector. The institution's embrace of modern banking technologies reflects its dedication to enhancing accessibility, efficiency, and security for clients, ultimately fostering a dynamic and resilient financial ecosystem.

#### **1.5 Statement of Problem**

The rapid digitalization of financial transactions has led to an increased vulnerability to fraudulent activities in banking systems. This problem of financial fraud in banking transactions is a serious and costly issue that affects both banks and their customers. In order to create a reliable model that

can identify and stop financial fraud in banking transactions, it is imperative to use cutting-edge technologies, particularly neural networks, as traditional fraud detection techniques are frequently unable to keep up with the volume and sophistication of fraudulent activities.

### **1.6 Research Objectives**

1. To develop and evaluate an artificial neural network model for fraud detection
2. To validate the model's performance using real world banking data and predict the fraudulent status of the next five customers.

### **1.7 Research Questions**

1. How can a neural network be developed and assessed to ensure its effectiveness in detecting fraudulent transactions, while also considering the feasibility of implementation by banks?
2. What is the performance of the developed neural network model when tested with real-world banking transaction data?

### **1.8. Scope of the Study**

The scope of this study is defined by the specific boundaries and parameters within which the research is conducted. The study is focused on application of neural networks for detecting and preventing financial fraud in banking sector, with a primary emphasis on the case study of the Central African Building Society (CABS). The scope of the study include following key elements:

#### **Geographical Scope:**

The primary focus is on the Central African Building Society, situated within its operational and regulatory environment. The findings and recommendations are primarily applicable to the context of this specific banking institution.

#### **Fraud Detection Methodology:**

The research concentrates on the use of neural networks as an advanced tool for fraud detection. The scope encompasses the specific neural network architecture chosen for the study, training procedures, and the evaluation of its effectiveness in identifying and preventing financial fraud.

#### **Case Study Focus:**

The study delves into the unique challenges and dynamics of the Central African Building Society as a case study. Insights derived from this case study are central to the research, but generalizations to other banking institutions are made with caution.

### **1.9 Limitations:**

The study explicitly recognizes certain limitations, such as case study specificity and data quality issues, which shape the boundaries of the research and influence the interpretation of findings. It is essential to be aware of these scope-related considerations when interpreting the study's results. While the findings contribute valuable insights to the field, they are contextualized within the defined scope, and generalizations should be made judiciously based on the study's specific parameters.

### **1.10 Significance of the Study**

This study's importance comes from its capacity to address critical challenges in the contemporary banking landscape. As the financial sector undergoes rapid technological transformations, the escalation of financial fraud poses a substantial threat. By focusing on application of neural networks on fraud detection, this study aims to contribute in the following ways:

### **1. Enhancing Security Measures:**

The research explores cutting-edge solutions to counteract the evolving nature of financial fraud. Neural networks offer the potential to enhance security measures, providing a proactive approach to effectively identify fraudulent activity.

### **2. Informing Banking Practices:**

The findings of this study, particularly through the case study of the Central African Building Society, can provide valuable insights for banking institutions. Practical implications derived from the research may inform and guide banking practices, aiding institutions in strengthening their defenses against fraud.

### **3. Contributing to Academic Knowledge:**

By conducting a comprehensive exploration of neural networks in the context of financial fraud, the study contributes to academic knowledge. It adds to the existing literature on the intersection of artificial intelligence and finance, offering a nuanced understanding of the role neural networks can play in bolstering security measures.

### **4. Guiding Future Research:**

The study paves way for future research to endeavors in financial fraud detection researchers. It may inspire further investigations into refining and optimizing neural network models, adapting them to diverse banking environments, and exploring novel applications in the broader context of financial security.

### **5. Promoting Ethical AI Implementation:**

As the study delves into the practical implementation of neural networks, it has the potential to foster discussions around ethical considerations in AI. By addressing issues related to transparency, accountability, and fairness, the study contributes to responsible AI implementation in banking departments.

### **1.11 Assumptions of the Study**

In conducting this study, certain assumptions are made to guide the research process and interpretation of findings. These assumptions provide a framework for the investigation, acknowledging certain conditions or premises that may influence the study outcomes. The key assumptions include:

#### **1. Data Accuracy and Integrity:**

It is assumed that the data collected for the study, both historical and contemporary, is accurate and reflects the true nature of the banking transactions and fraud instances. Any discrepancies or inaccuracies in the data could impact the reliability of the study

#### **2. Representativeness of the Case Study:**

The assumption is made that the Central African Building Society (CABS) serves as a representative case study for the broader banking sector. Findings and insights derived from the CABS case study are considered applicable to some extent to similar banking institutions facing comparable challenges.

#### **3. Neural Network Model Effectiveness:**

The study assumes that the chosen neural network model, as detailed in the methodology, is effective in detecting and preventing financial fraud. The success of the neural network is contingent on factors such as appropriate training, relevant features, and the quality of the data.

### **1.13 Definition of Terms**

#### **Neural network:**

It is a series of algorithms that endeavors to recognize underlying relationships in a set of data through a process that mimics the way the human brain operates. Neural networks refer to systems of neurons either organic or artificial in nature. It is a set of algorithms designed to replicate the way the human brain functions in order to identify underlying links within a piece of data. Artificial or natural networks of neurons are referred to as neural networks, (He, 2023)

**Financial fraud:**

It refers to any intentional act or omission that misleads others in order to obtain financial gain or cause financial loss. It is a deceptive and unlawful activity with the intent of gaining unauthorized access to financial resources or manipulating financial transactions, (YA., 2020)

**Fraud detection:**

According to (Wang, 2019) fraud detection is a process of detecting and stopping fraudulent activities.

**Network architecture:**

Network architecture refers to the structured design of a neural network, encompassing its layers, (LeCun, 2015)

**Security Measures:**

Security measures are strategies and protocols designed to protect financial transactions and data to unauthorized access, breaches, and other security threats. (Sun, 2022)

**1.14 Chapter Summary**

In conclusion, this introductory chapter serves as the foundation framework for the comprehensive exploration into the application of neural networks for fraud detection in banking sector, a case study of the Central African Building Society (CABS). In laying this groundwork, Chapter1 establishes the context, rationale and structure for the subsequent chapters. The upcoming chapters delves into the theoretical foundations, research methodology, case study analysis and the synthesis of findings

## **CHAPTER 2:**

### **LITERATURE REVIEW**

#### **2.1 Introduction**

This research is based on a number of previous studies that have been done on Financial Fraud and Neural networks. The literature review offers a thorough analysis of existing knowledge regarding the utilization of neural networks to financial fraud detection within the banking industry. This dissertation's theoretical framework draws on several key frameworks and concepts that underpin the research on financial fraud detection, neural networks, and their integration within the banking industry. This chapter aims to establish a theoretical basis for the subsequent exploration of the use of neural networks to stop financial fraud, with a specific focus on the Central African Building Society (CABS). The chapter aims to synthesize theoretical and empirical literature, identifying key concepts, methodologies, and gaps that pave the way for the subsequent chapters in this dissertation.

#### **2.2 Theoretical literature**

##### **2.2.1 Fraud detection**

Financial Fraud detection refers to the process of figuring out if a transaction is fraudulent or not, (A. S. Kumar et al, 2023). Wang, (2019) defined it as the process of spotting and preventing fraud. It is a critical challenge faced by financial institutions and consumers alike. Dr. V. Ragavarthini et al, (2024) states that sophisticated analytics, machine learning algorithms, and artificial intelligence are used in financial transaction fraud detection to identify anomalous behaviors of fraudulent activities. By analyzing enormous volumes of transactional data in real-time, financial institutions can swiftly detect or mitigate fraudulent activities, thus safeguarding assets, maintaining trust and preserving financial integrity. The detection of fraud entails monitoring the activities of the user population in order to estimate, identify or prevent objectionable behavior, which consist of fraud, intrusion and defaulting, (S.P. Manraj, 2019).

##### **2.2.2 Types of Financial Fraud in banking transactions**

Approximately 71% of occurrences of online payment fraud are the result of identity theft. In some cases, fraudsters steal personal information to access bank accounts or open new accounts in someone else's name, obtain card information, and use the stolen information to make online transactions. The purchased items are then delivered to the fraudster, while the legitimate cardholder can start a chargeback procedure often incurring associated fees, (R. V. Patricia 2023).

When an unauthorized individual makes purchases or withdraws money using someone else's credit card information, it is known as credit card fraud. Types of credit card fraud are Skimming, which Involves capturing card information using a device attached to an ATM or POS terminal. Phishing in which Fraudsters trick individuals into revealing their credit card details through fake websites or emails and card-not-present (CNP) Fraud in which Fraudulent transactions are conducted online or over the phone where the physical card is not required, (Jurgovsky, J., Granitzer, M., Ziegler, K., et al. 2018).

According to McNeil, D. M., & Michael, M. (2019), account takeover fraud occurs when a fraudster obtains an unauthorized access to a victim's bank account and conducts unauthorized transactions, often changing account details to lock out the legitimate account holder. This includes Credential Stuffing where fraudsters use automated tools to try multiple username-password combinations and Social Engineering where fraudsters Manipulate individuals into revealing confidential information.

Wire transfer fraud is another type of financial fraud in banking transactions which involves illegally transferring money from a victim's account to another account controlled by the fraudster. Examples are Business Email Compromise (BEC) where fraudsters compromise business email accounts to initiate unauthorized wire transfers and Fake Invoices which involves sending fake invoices to businesses to trick them into transferring funds to the fraudster's account, (Abdallah, A., Maarof, M. A., & Zainal, A. 2016).

Another type of financial fraud is Check fraud. It involves the use of counterfeit or altered checks to illegally acquire funds from a bank account. This involves forgery which is altering the details on a check to change the payee or amount and counterfeit checks where fraudsters create fake checks that appear legitimate, (Nielsen, K. B. 2020).

Kroll. (2020), stated that insider fraud occurs when employees of a bank or financial institution commit fraud using their access to sensitive information and systems. Insider trading involves Embezzlement and Unauthorized Transactions. Embezzlement refers to the misappropriation of funds by employees and Unauthorized Transactions involves conducting transactions without the customer's knowledge or consent.



According to Symantec. (2019), online banking fraud involves unauthorized access to online banking accounts, leading to unauthorized transactions. Phishing and Vishing and Malware are types of online banking fraud. Phishing and vishing refer to Scamming individuals through fraudulent emails and calls to gain login credentials malware is the use of malicious software to capture login information or redirect transactions.

These types of fraud can overlap and often use similar methods, making it critical for financial institutions to employ robust detection and prevention strategies. Using neural network classifiers can improve the capacity to identify and effectively address these fraudulent schemes.

### **2.2.3 Traditional Fraud Detection Techniques**

Historically, rule-based structures and statistical models have historically formed the mainstay of the detection of fraud. While rule-based systems set predefined criteria, statistical models rely on deviations from established norms. Research by Johnson and Smith (2015) highlights the limitations of these methods, emphasizing their challenges in adapting to the constantly evolving strategies of fraud perpetrators.

Classical fraud detection uses rules solely to prevent fraudulent transactions, however, this approach produces a large number of false positives since it blocks legitimate customers if, for example a limit is defined regarding when to disable.

Before the emergency of neural networks or any machine learning algorithms, credit card detection exclusively depended on rule-based systems, manual verification procedures and a variate of authentication techniques, (Nguyen, 2020).

Based on the pre-established rules, rule-based systems have been used to detect potentially fraudulent transactions. These regulations may be as basic as noting transactions over a specific threshold or as intricate as spotting odd transaction trends. The systems use an array of logic-based rules to classify dubious and non-suspicious activities, for example, if a several transactions happen quickly one after the other, this may be reported as suspicious.

Credit card fraud can also be detected using manual method. The manual method entails human analysis and transactions verification. For instance, the credit card company may get in touch with the customer directly to confirm a transaction if it appears unusual given their spending patterns.

In some instances, transactions that are made at a location other than the customer's typical location may also be flagged for manual verification.

Additionally, procedures for authentication have been put in place. These precautions include asking extra information during a transaction like a zip code or a PIN code. The foundation of this technique is multifactor authentication, which entails confirming the cardholder's identity with a number of different sources of information or factors. According to Nguyen (2020), these variables may consist of something the user knows (like a password), something they possess (like a physical card), or something they are (like fingerprint).

These methods continue to be applicable in modern contexts and have demonstrated their efficacy in the past, usually in conjunction with more advanced machine learning techniques. However, they do have some limitations, like the need for human intervention and the difficulty of keeping up with ever-changing fraudulent schemes. As a result, machine learning algorithms have become increasingly important tools in the fight against credit card fraud, mainly because of their ability to automatically identify and adapt to new fraudulent behaviors, (Kulatileke, 2020)

#### **2.2.4 Limitations**

Though they work well in many situations, traditional fraud detection techniques have a number of drawbacks. Rule-based systems need to be updated often to stay up to date with the constantly changing fraud patterns. Because these criteria are manually created, their capacity to adjust to new fraud categories or slight modifications in fraud trends is limited. The technologies have the potential to produce a large number of false positives, which could result in needless verification steps and possibly unhappy customers. As more rules are added, rule-based systems may become inefficient since each transaction must be compared to every rule.

Additionally time- and labor-intensive are manual verification techniques. They may also be inconsistent since they rely on the judgment and availability of human operators. Customers may find the human verification processes bothersome or invasive, particularly if they are regularly contacted to confirm transactions.

#### **2.2.5 Emergence of Neural Networks**

Development on neural networks got underway in the early 1940s. Its popularity increased dramatically in the late 1980s as a result of the development of new methods and strategies as well as broader developments in computer hardware technology.

While some neural networks are biological neural networks' models and some are not, historically, the goal of creating artificial systems that could perform complex, potentially "intelligent," computations akin to those the human brain does on a daily basis served as a major source of inspiration for the field of neural networks and may have improved our understanding of the human brain. The emergence of neural networks, motivated by the human brain's learning mechanisms, offers a paradigm shift in fraud detection. Deep learning systems, like CNNs and RNNs, exhibit exceptional capabilities in discerning intricate patterns from data, (Goodfellow, Bengio, & Courville, 2016).

The research conducted by Mitchell and Anderson (2017) and Chen et al. (2019) underscores the transformative potential of neural networks in addressing the shortcomings of traditional methodologies. The theoretical underpinning of neural networks is based on the idea of adaptive learning. Neural networks, inspired by the human brain, may learn from patterns in data and modify their parameters to optimize performance. The work of Rumelhart et al. (1986) on backpropagation introduced the concept of iterative learning, providing a theoretical foundation for the adaptability of neural networks in recognizing evolving patterns associated with financial fraud, (Chen, Yan, & Wang, 2019; Mitchell & Anderson, 2017).

Contemporary research highlights the effectiveness of neural networks in fraud detection by leveraging their ability to process and examine large datasets, looking for trends and anomalies that points to fraudulent activities. Neural networks, particularly deep learning models, have demonstrated superior performance in detecting complex fraud patterns compared to traditional methods (Alazab et al., 2021; Randhawa et al., 2018). These models are grounded in decision theory principles, enhancing their capability to operate under uncertainty and improve decision-making accuracy in fraud detection scenarios (Ngai, Hu, Wong, Chen, & Sun, 2011).

Furthermore, advancements in explainable AI have addressed some of the challenges related to the interpretability of neural network decisions, aligning with the bounded rationality framework by providing more transparent decision-making processes (Guidotti et al., 2019). This development ensures that neural network-based fraud detection systems are not only effective but also comprehensible to human analysts, facilitating better decision-making and trust in automated systems.

### **2.2.6 Neural networks in fraud detection**

Frauds in banking transactions are common these days as most of the people worldwide now use the credit card payment methods regularly. This is due to technological advancement and a growing amount online transactions resulting in frauds causing enormous financial losses.

As technology advanced, so did fraudsters' tactics hence neural networks begun to be applied in fraud detection systems around late 1990s to early 2000s. Researchers and practitioners started experimenting with using neural network models to identify patterns of fraudulent behavior in banking transactions, like credit card and payment fraud. These artificial intelligence-driven approaches delved deep into historical data, uncovering intricate fraud patterns that eluded rule engines, (Y. Patel, 2023). Neural networks gained traction in the fraud detection space as a result of their ability to detect complex, non-linear data patterns. Financial institutions and payment processors started incorporating neural network-based models into their fraud detection systems to improve accuracy and efficiency in identifying suspicious activities.

Neural network adoption in fraud detection applications increased with advances in computer power and large-scale dataset availability. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs), two deep learning techniques, have been used to evaluate transactional data and identify fraudulent patterns more accurately. The Awoyem group (2017)

Neural networks, particularly deep learning models, have become integral components of modern fraud detection systems. Organizations across different industries, such as finance, e-commerce, healthcare, and telecommunications, rely on neural network-based models to detect and prevent fraudulent activities in real-time, (M. L. Smith 2014)

Throughout these periods from the late 1990s, neural networks have continually evolved, incorporating advancements in architecture, training algorithms, and data processing techniques to enhance their effectiveness in fraud detection. Today, neural network-based fraud detection systems support sophisticated deep learning models to analyze vast amounts of transactional data and identify fraudulent behaviors with high precision and efficiency.

### **2.2.7 Population size**

The population size has a great impact on the neural network algorithm performance. The algorithm will not be able to produce correct results if the population size is too small. Cross-

validation and grid search among several others are techniques that can be used to find optimal population size, (Zhang et al 2016).

In machine learning, grid search is a method for rotating hyperparameters to determine the ideal set for a particular model. It operates by repeatedly going through a range of potential values for every hyper-parameter and assessing the model's performance for every combination. The ideal collection of hyper-parameters is then determined by combining the values that result in the highest performance. Although grid search is a straightforward and simple approach to use, it can be computationally expensive. (Ludecke 2019).

Techniques like sample splitting and resampling that use different data segments to test and train a model on consecutive iterations are referred to as cross-validation. It is widely used when the primary goal is to anticipate outcomes and one wants to determine how well a predictive model would perform in actual settings. details. The data ought to be split into training and validation sets, basically. The validation set is used to evaluate the model's performance on new data after it has been trained on the training set. According to Srivatsa (2022), "k-fold cross validation" is the most often utilized technique for data splitting, despite the fact there are other techniques.

#### **2.2.8 Theoretical review of fraud detection techniques and algorithms**

Algorithms for detecting fraud make use of artificial intelligence and machine learning to assess data and spot fraudulent activity. Essentially, detecting fraud is a classification task that can be tackled by supervised, reinforced, or unsupervised techniques.

#### **2.2.9 Neural network classifier**

Because neural networks can represent intricate, non-linear interactions, they have become more and more popular, particularly deep learning models. They are made up of numerous layers of networked neurons that can learn to represent the data hierarchically. Typically, neural networks are trained using the backpropagation algorithm, which also modifies weights to reduce prediction error. (Hinton, LeCun, and Bengio 2015).

#### **2.2.10 Types of neural network classifiers**

The most basic kind of artificial neural network is called a feedforward neural network (FNN), in which node connections do not form cycles. Because of their simple architecture, these networks are commonly used for tasks like binary and multi-class classification. In a conventional FNN, data flows in one direction—from the input layer, through hidden levels, to the output layer—without any feedback loops. This makes FNNs useful for jobs where the input-output relationship

may be trained without needing to account for temporal dependencies or sequences within the data, (Nielsen, 2019; Goodfellow et al. 2016).

Convolutional neural networks, or CNNs, are especially made to handle input that has a structure resembling a grid, like photographs. They use convolutional layers, which apply different filters to the input data, to capture spatial hierarchies and patterns very effectively. CNNs have been modified to evaluate transaction data as a type of temporal or spatial data in the context of fraud detection, allowing them to identify complex patterns suggestive of fraudulent activity. According to (Zhou et al. 2023; Fu et al. 2016), CNNs can recognize these patterns by converting transaction sequences into feature matrices that convolutional layers can process efficiently.

Autoencoders are used for anomaly detection by learning a compressed representation of normal transaction data and identifying deviations from this norm, (Fiore et al. 2019) employed autoencoders in fraud detection, achieving significant improvements in detecting outliers indicative of fraud.

Because they can keep track of past inputs, Recurrent Neural Networks (RNNs) are especially useful for processing sequential data. This is especially useful for jobs like time-series analysis and transaction sequence monitoring. Because of this feature, RNNs can use previous data to affect the current output, which makes them useful for a variety of tasks like financial fraud detection, speech recognition, and natural language processing, (Encord, 2022).

Long Short-Term Memory (LSTM) networks, a specialized type of RNN, enhance this capability by addressing the limitations of traditional RNNs, such as the vanishing gradient problem. LSTMs can capture long-term dependencies within data sequences through their unique architecture, which includes memory cells and gating mechanisms. These features enable LSTMs to retain information over extended periods and effectively model time-series data and transaction sequences, which are essential for detecting patterns indicative of fraudulent activity (Walker et al., 2022; Built In, 2023).

LSTMs have been widely adopted in various domains for tasks requiring long-term memory. For instance, in financial fraud detection, LSTMs can analyze sequences of transactions to identify anomalies that deviate from typical behavior patterns, thereby providing a robust mechanism for predicting and preventing fraudulent activities, (Abbass et al., 2021; Greff et al., 2019). The architecture of LSTMs, with their ability to selectively remember and forget information, makes them particularly effective for these applications, (Haider et al., 2022).

In summary, the integration of RNNs and LSTMs in financial fraud detection leverages their sequential data processing capabilities and long-term memory retention to identify and prevent fraudulent activities effectively. These neural network models, by capturing temporal dependencies and patterns, provide significant advantages in the real-time monitoring and analysis of transaction sequences, (Hochreiter & Schmidhuber, 1997; Encord, 2022; Walker et al., 2022).

## **2.3 Empirical literature**

### **2.3.1 Related studies**

Over time, numerous machine learning algorithms have been put forth to identify and stop financial fraud in banking transactions, according to (Yang et al., 2020). The authors of the study Credit Card Transaction Fraud Detection Using Neural Network Classifiers, E. Nazeriha, (2023), created synthetic samples for the minority class using the Synthetic Minority oversampling Technique, Variational Auto-encoders, and Generative Adversarial Network in order to create a more balanced dataset.

The study's findings showed that GAN does not perform better than the other classifiers because, in three of the five classifiers, the generated samples from VAE were the most successful. Doing dataset is one way to make things work better. Correlation analysis, for example, can be used to identify the characteristics that have the biggest influence on classification. Then, the most significant features can be dropped, and the trimmed-down samples can be used to train classifiers and generative algorithms.

In their research A Neural Survival Analysis For Fraud Early Detection, Panpan Zheng et al. (2018) introduced the SAFE fraud early detection model, which is based on survival analysis and maps dynamic user activities to probability of survival that are guaranteed to decrease monotonically with time. Recurrent neural networks (RNNs) are used by SAFE to process user activity sequences and provide dangers at each time stamp. Consistent predictions are then made by applying survival probability, which is derived from hazard values. The results on two real-world datasets show that SAFE performs better than state-of-the-art fraud early detection techniques, the survival analysis model, and the recurrent neural network model alone.

The objective of the 2022 study Fraudulent financial transactions Detection Using Machine Learning by Mosa. M. M. Megdad was to evaluate multiple machine learning algorithms to

ascertain the accuracy and speed at which financial transactions may be identified as authentic. Among the techniques used by the researcher were the MLP repressor, Random Forest Classifier, K Neighbors Classifier, Logistic Regression, Bagging Classifier, Decision Tree Classifier, and Deep Learning. Information was obtained from the Kaggle database. On an unbalanced dataset, the Random Forest Classifier performed best with 99.96% precision, 99.97% accuracy, 99.97% recall, and 99.96% F1-score. However, the best classifier with a balanced dataset was the Bagging Classifier, with 99.95% precision, 99.96% accuracy, 99.98% recall, and 99.96% F1-score.

The use of neural networks and logistic regression for fraud detection in financial transactions was suggested in a different paper, Neural Networks for Fraud Detection in Financial Transactions by Abhiraj Kulkarni, (2023). The researcher created a machine learning model that accounts for the financial transactions' time-sensitive nature as well as the fraud detection process' inherent imbalance. The suggested model extracted features from the transactional data and predicted the probability of fraud using a combination of long short-term memory networks and convolutional neural networks.

The suggested model works better than the current approaches, as evidenced by the results, which also show a higher F1 score and area under the receiver operating characteristic curve.

## **2.4 Research Gap**

While existing literature has significantly advanced our understanding of the application of neural networks in financial fraud detection, several research gaps and areas for further exploration emerge. Identifying and addressing these gaps is essential for advancing the field and providing meaningful insights for the case study of the Central African Building Society (CABS).

### **1. Customer Perception and Trust**

The empirical literature briefly touches on customer perception, but a more in-depth exploration of how customers perceive and trust neural network-based fraud detection systems is warranted. Understanding customer attitudes and concerns is crucial for successful implementation, and this area remains underexplored in the current body of literature.

### **2. Long-Term Effectiveness and Adaptability**



While some studies hint at the adaptive learning capabilities of neural networks, there is a gap in long-term analyses of their effectiveness and adaptability to continually evolving fraud tactics. A more extensive investigation into the sustained performance of neural networks over time and their ability to address emerging threats is needed.

### **3. Integration of Regulatory Compliance and Fraud Detection**

The literature acknowledges the intersection of technology, regulatory compliance, and fraud prevention, but there is a research gap in understanding the intricate dynamics of this integration. A more thorough exploration of how neural networks align with and contribute to regulatory compliance requirements within the banking sector is necessary.

### **4. Comparative Analysis Across Regional Banks**

While individual case studies exist, a broader research gap lies in a comprehensive comparative analysis across regional banks facing similar challenges. A comparative study could reveal nuanced differences and commonalities in the implementation and outcomes of neural network-based fraud detection systems, providing valuable insights for institutions like CABS.

### **5. Impact of Socio-Economic Factors on Fraud Dynamics**

The influence of socio-economic factors on fraud dynamics is briefly mentioned in the literature, but a research gap exists in a detailed examination of how these factors may shape fraud patterns in the Central African context. Exploring the socio-economic landscape can contribute to a more holistic understanding of fraud risks.

## **2.5 Synthesis of Research Gaps**

Addressing these research gaps is imperative for advancing the field of neural network-based fraud detection in banking, especially within the Central African context. The dissertation will strive to fill these gaps by conducting a detailed case study at CABS, providing nuanced insights into the practical application and outcomes of neural networks for fraud prevention in this specific environment.

## **2.6 Proposed Conceptual Model**

A conceptual framework, according to Varpio (2020), is a figure that illustrates the anticipated relationship between the cause-and-effect elements. A conceptual framework elucidates the relationships between the variables through a series of related terminology, definitions, and

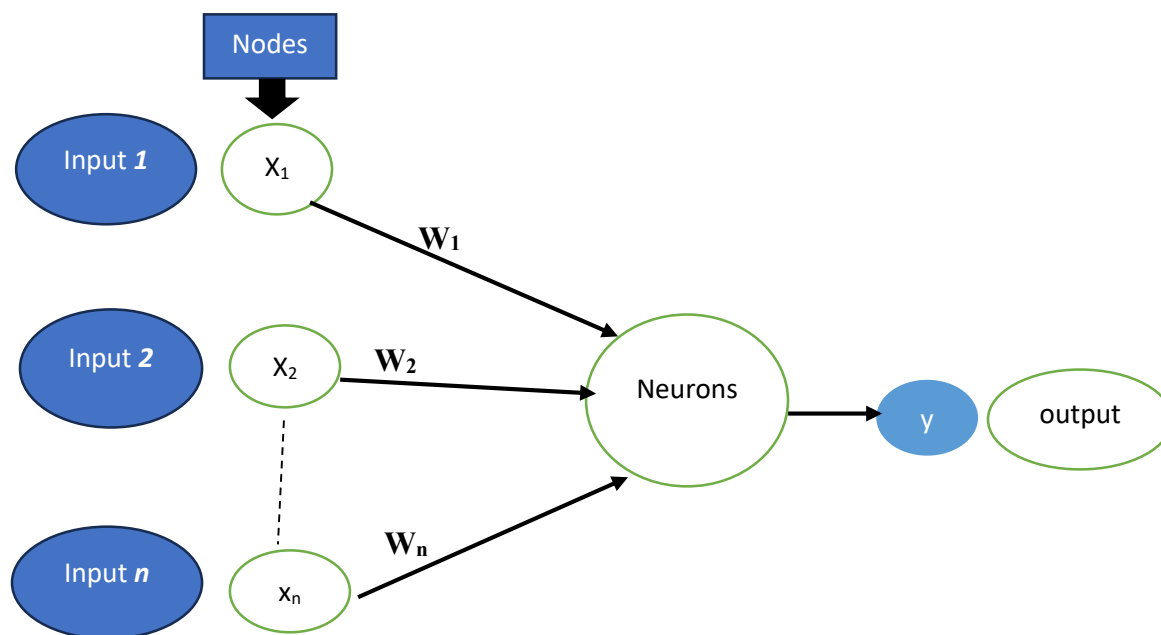
assertions. A variable is an aspect of a phenomenon that is observable or quantifiable. (Collis and Hussey 2009).

The model outlines key variables, hypotheses, and moderating variables. This model serves as a roadmap for the empirical investigation within the specific context of the Central African Building Society (CABS).

Artificial Neural Network Classifiers are used by the proposed system to distinguish between fraudulent and valid transactions. Accuracy is computed via prediction, and performance is measured.

**Figure below shows a proposed conceptual model**

*Figure 2.1 proposed conceptual model*



An artificial neural network's biological model is the human brain. Neurones in the human brain are interconnected similarly to nodes in an artificial neural network. The above figure depicts the input, output, and hidden layer structure of an ANN classifier.  $X_1, X_2, \dots, X_n$  are the inputs, and  $y$  is the output.  $W_1, W_2, \dots, W_n$  are the weights that correspond to the inputs  $X_1-X_n$ .

## **2.7 Advantages of neural networks**

Because deep learning techniques, particularly neural networks, can automatically learn complicated patterns and representations from raw data, they present a promising alternative for the identification of financial fraud, (Takács & Barna, 2021). This removes the requirement for labor-intensive, domain-specific manual feature engineering. According to Asim et al. (2022), it is a faster and more accurate way of detecting fraud that greatly reduces the possibility of bias or human mistake affecting the data needed for fraud detection.

Artificial neural networks contribute to the accurate detection of fraud at the same moment it occurs, ideally prior to the fraudulent transaction being processed, (Takács & Barna, 2021). Artificial neural networks, for instance, have a 95% accuracy rate when it comes to detecting credit card fraud, according to Human-Centric Intelligent Systems. Artificial neural networks are also a more effective tactic because fraud prevention involves more than just identifying and thwarting threats at the outset. Business leaders can detect flaws in their systems and be better equipped to address them with the use of machine learning methods like artificial neural networks and thorough research of key performance metrics linked to potentially fraudulent transactions or claims. Additionally, this makes risk analysis possible, which aids firms in avoiding possible fraud sources such people with questionable credit, personnel with questionable backgrounds, companies making outrageously good-looking claims, and so forth (Asim et al., 2022).

## **2.8 Limitations**

A significant amount of high-quality data is required for machine learning algorithms to perform properly. The accuracy of the algorithm's predictions could be compromised by biased or insufficient training data. It can be challenging to analyze and comprehend artificial neural networks, particularly for those who are not familiar with the technical specifics of how they operate. People may find it challenging to comprehend why specific transactions are being flagged by the system as possibly fraudulent as a result. Furthermore, machine learning algorithms can be costly to set up and keep up, particularly if a business lacks internal expertise in this field. In addition, human intelligence is lacking. When it comes to analyzing and interpreting data to assess the likelihood of dubious activity, even the most sophisticated technology cannot fully replace the knowledge and discretion of a human. In order to properly filter and evaluate data in order to

ascertain the significance of a risk score, human psychological analysis and comprehension are essential, (Sanghvi, H. 2023).

## **2.9 Chapter summary**

In conclusion, the extensive exploration of theoretical frameworks and empirical studies on the application of neural networks in financial fraud detection provides a robust foundation for this dissertation. Theoretical frameworks such as the Fraud Triangle, adaptive learning principles, and decision theory offer valuable lenses for understanding the complexities of fraud dynamics and the functioning of neural networks. The literature review synthesizes theoretical and empirical insights, identifies research gaps, and introduces the proposed conceptual model. It sets the stage for the subsequent chapters, emphasizing the significance of investigating neural network applications in financial fraud detection, particularly within the unique context of CABS in Central Africa.

Empirical studies present compelling evidence of the efficacy of neural networks in enhancing fraud detection within the banking sector. These studies reveal improvements in accuracy, reductions in false positives, and adaptability to evolving fraud patterns. However, the current literature has certain notable gaps that underscore the need for further research, particularly in the specific context of the Central African Building Society (CABS).

The Identified research gaps, including the lack of context-specific applications, customer perception, and a comprehensive comparative analysis across regional banks, present opportunities for this dissertation to contribute significantly to the field. The empirical literature hints at the promising potential of neural networks but leaves room for a nuanced examination within the unique socio-economic, regulatory, and technological landscape of Central Africa.

As we delve into the case study of CABS, these research gaps will guide our investigation, aiming to provide not only practical insights for the specific organization but also contributing to the broader academic discourse on the application of neural networks for financial fraud detection. The empirical findings from the case study will be instrumental in refining and validating the theoretical concepts discussed in this literature review.

In the subsequent chapters, we will transition from the comprehensive literature review to the methodology employed for this research. The methodologies will be designed to address the

identified gaps and contribute new knowledge to the field of neural network applications in the banking sector, specifically focusing on CABS in the Central African context.

## **CHAPTER 3:**

### **RESEARCH METHODOLOGY**

#### **3.1 Introduction**

This chapter describes the approach taken by the researcher to look into credit card fraud detection and prevention using neural networks. The study's design, data sources, target population, sampling techniques, instruments, data collection strategies, variables, data processing techniques, ethical considerations, and anticipated results are covered in the conclusion.

#### **3.2 Research Design**

A combination of methods and procedures known as a research design are used in the measurement and analysis of the variables specified in the problem research. The research problem, hypotheses, independent and dependent variables, experimental design, and, if applicable, methods for gathering data and a strategy for statistical analysis are all identified in a study's design. Descriptive-longitudinal case studies are one of the study types and subtypes that include descriptive, correlational, semi-experimental, experimental, review, and meta-analytic research. A research design is a framework created to address research issues (Sagar Mozumder, 2021).

The research follows a quantitative approach, utilizing a combination of descriptive and predictive analytics. It involves the development and implementation of neural network models to detect and prevent credit card fraud.

Exploratory research is a study that seeks to answer a question or address a phenomenon, (Hair et al, 2019). In the context of financial fraud detection, exploratory research can be used to uncover hidden patterns or relationships in the data that may not be immediately apparent. This type of research involves conducting preliminary investigations, such as data visualization, clustering analysis, or anomaly detection, to identify interesting trends or anomalies in the data. Exploratory research can help in generating hypotheses and guiding further investigation into the underlying causes of fraudulent behavior.

#### **3.2 Data Sources**

A data source is the initial location where the data is born. Data for this study is sourced from CABS Bank's transaction records, encompassing both legitimate and fraudulent credit card transactions. Additional sources include historical fraud data, customer information, and transaction metadata.

### **3.3 Target Population and Sampling Procedures**

The users of credit cards and transactions made at CABS Bank in Zimbabwe make up the target population. The fact that CABS Bank Zimbabwe is the biggest bank in Central and Southern Africa had a role in the decision. Given that CABS is among the banks with the greatest number of active clients in the banking sector, there is a higher likelihood of fraudulent transactions.

The investigator employed a deliberate sampling approach in order to categorize the sample units. A non-probability sample known as a judging or expert sample is a type of deliberate sample (Tallam, 2016). The main goal of a purposeful sample is to provide a sample that can be credibly held to be representative of the entire population. This is also accomplished by using the expertise of the population to select a non-random sample of the components that make up a cross-section.

The investigator utilized a purposeful sampling technique to classify the sample units. One kind of non-probability sample is purposeful sampling, sometimes referred to as expert or evaluating sampling (Lavrakas, 2008). The primary objective of a purposeful sample is to produce a sample that can be reasonably expected to be representative of the total population. This is also achieved by employing the population's expertise to choose a non-random sample of the constituents of a cross-section.

### **3.4 Methods for Data Collection**

Data collection methods include automated transaction monitoring systems, manual verification processes, and periodic audits. Additionally, customer feedback surveys and interviews with bank staff may provide qualitative insights into fraud detection and prevention practices.

### **3.5 Variables and Expected Relationships**

Variables include transaction amount, location, time, frequency, customer demographics, and transaction metadata. The researcher focused on two variables which are transaction amount and time. The expected relationship is that fraudulent transactions will exhibit anomalous patterns compared to legitimate ones, allowing the neural network to learn and detect fraud effectively.

### **3.6 Diagnostic Tests**

#### **3.6.1 Data quality checks**

Ensuring the accuracy and completeness of data is crucial, as the input directly impacts output quality. Before any data analysis the data will undergo rigorous checks to ensure completeness and

accurate. Missing information will be handled through imputation techniques while outliers will be treated or removed based on predefined criteria to prevent skewed results.

### **3.6.2 Feature correlation analysis**

This involve examining the relationships between various features using the spearman rank correlation, the goal is to identify features that provide unique information and eliminate those that may introduce redundancy into the model, potentially skewing

### **3.6.3 Multicollinearity analysis**

Variance Factor Analysis was used to measures the correlation between one independent variable and the others in a regression model

## **3.7 Analytical Model**

The neural network that is used is a feed forward neural network, often known as a multilayer perceptron (MLP). This type of neural network consists of an input layer, one or more hidden layers, and an output layer. A detailed description of the architecture and its components may be found below:

## **3.8 Neural network architecture**

### **Input Layer:**

The input layer takes the preprocessed transaction data.

The quantity of characteristics in the input data (the training data's form) corresponds to the number of neurons in the input layer.

### **Hidden Layers:**

#### **First Hidden Layer:**

Contains 64 neurons.

Uses the ReLU (Rectified Linear Unit) activation function.

#### **Second Hidden Layer:**

Contains 32 neurons.

Also uses the ReLU activation function.

The hidden layers help the model learn complex patterns in the data by applying non-linear transformations.



**Output Layer:**

Contains a single neuron.

Uses the sigmoid activation function to output a probability score between 0 and 1.

This probability score indicates the likelihood of the transaction being fraudulent.

**Compiling the Model**

**Optimizer:** The model uses the Adam optimizer, which is an efficient optimization algorithm that adjusts the learning rate during training.

**Loss Function:** The binary crossentropy loss function is used because the problem is a binary classification problem (fraudulent vs. non-fraudulent transactions).

**Metrics:** The accuracy metric is used to evaluate the model's performance.

**Training the Model**

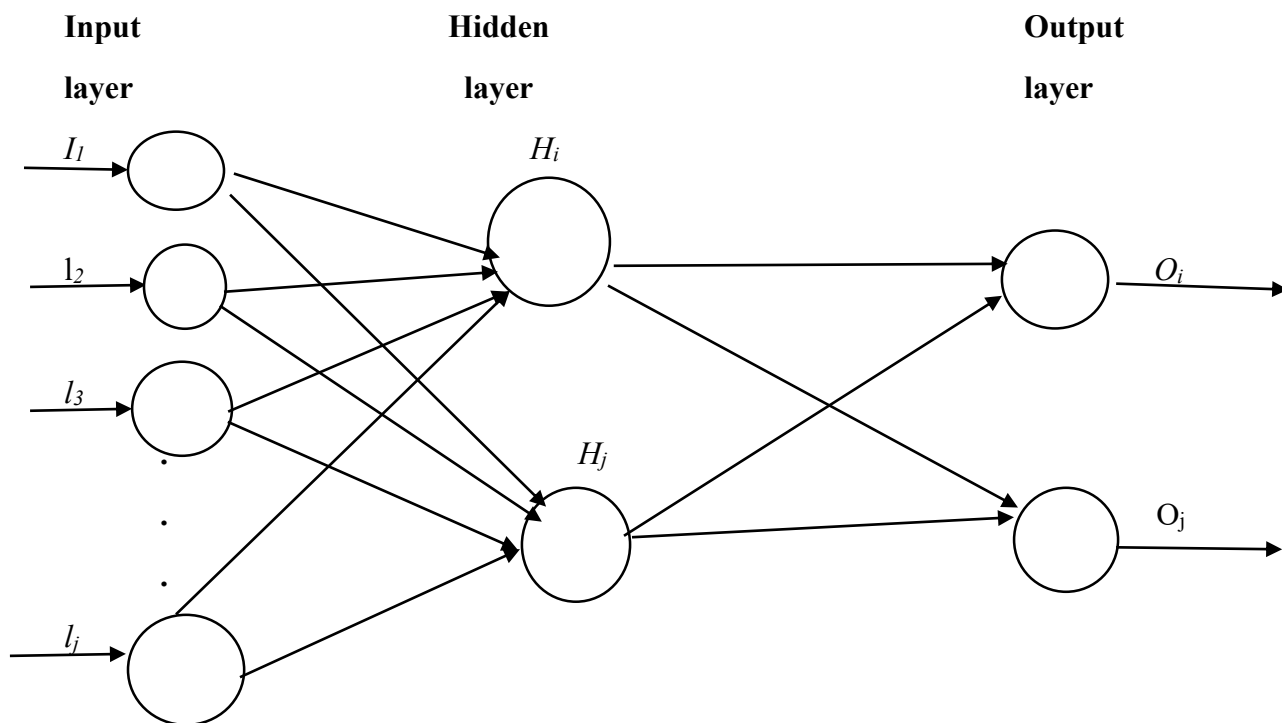
The model is trained on the resampled training data (with class imbalance handled using SMOTE).

It trains for 20 epochs with a batch size of 64.

A validation split of 20% is used to monitor the model's performance on a hold-out validation set during training.

**Figure below shows the neural network architecture**

*Figure 3.1 neural network architecture*



### 3.9 Model Validation (Fitness) Tests

The process of determining whether a selected statistical model is appropriate or not is known as model validation. In statistical inference, it is common for conclusions drawn from models that seem to suit the data to be anomalies, leading to researchers misjudging the true significance of their model. Model validation is used to determine whether a statistical model can withstand data variations in order to counteract this.

In order to make sure that a neural network model functions effectively on both newly discovered data and the training set, validation is an essential first step. This aids in evaluating the model's capacity for generalization.

### 3.10 Cross-Validation

The 'k'-fold cross-validation technique was applied to separate the data into k subsets. The remaining set was validated while the model was trained on 'k-1' subsets. In order to make sure the model functions well across various data sets, the process is repeated until each subset has been validated.

### **3.11 Ethical Considerations**

Safeguarding client confidentiality and privacy, maintaining data security, and getting informed consent for data usage are all ethical considerations. adherence to pertinent regulations and ethical guidelines for research involving human subjects are paramount. Ensuring that data used in research has been collected with the consent of individuals where necessary. De-identifying the data to protect individual privacy before analysis is also important. Lastly, adhering to regional and international data protection laws was observed during the research.

### **3.12 Chapter summary**

In conclusion, this chapter outlines the methodology adopted to investigate credit card fraud detection and prevention using neural networks at CABS Bank. By employing a quantitative research design, utilizing various data sources and collection methods, and ensuring ethical considerations, the study aims to develop effective strategies for mitigating fraud risks and enhancing financial security for bank customers.

## CHAPTER 4:

### DATA PRESENTATION, ANALYSIS AND INTERPRETATION

#### 4.1 Introduction

This chapter presents the findings and analysis derived from the study's investigation into the application of neural networks for detecting and preventing financial fraud in banking transactions, with a specific focus on the case study of the Central African Building Society (CABS). The chapter consolidates the outcomes of the conducted experiments and analyses, which are presented through descriptive statistics, model outputs, validation tests, and discussions of findings.

In this chapter, the researcher discusses and analyzes the results of the study using visual aids like tables, graphs, and charts within Jupyter Notebook. The researcher also compares the results from using neural network models with traditional methods like logistic regression to assess their efficacy in detecting fraudulent activities. Additionally, the chapter explores the insights gained from the neural network models and their implications for fraud detection and prevention. Various strategies to mitigate financial fraud in banks, informed by the study's findings, are also examined.

#### 4.2 Descriptive statistics

To provide a concise overview of the fundamentals and characteristics of the variables being studied, a preliminary analysis of the data used was carried out.

*Table 4.1 Descriptive statistics.*

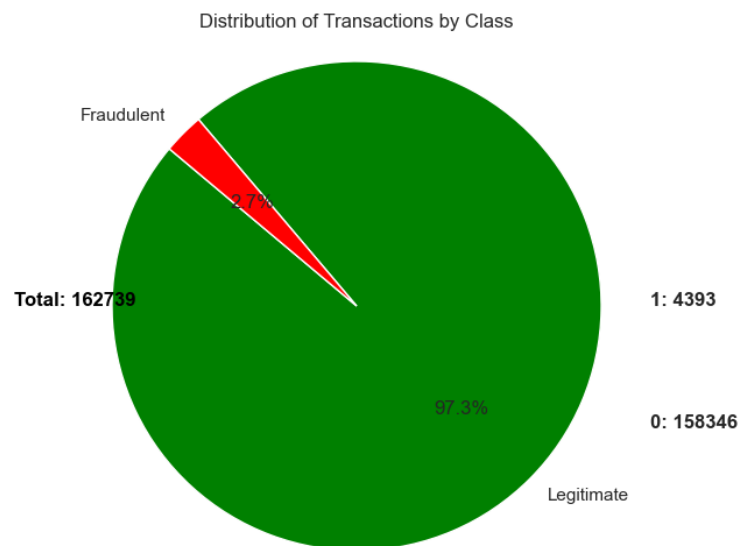
Variable	Count	Mean	Std	min	25%	50%	75%	Max	Skewness	kurtosis
Amount	162739.00	87.03	244.43	0.00	5.37	21.29	76.13	19656.53	15.75	678.70
Time	162739.00	58264.00	25414.02	0.00	40353.00	58906.00	76420.00	170348.00	-0.04	-0.18

The table 4.1 shows the important summary of descriptive statistics of the predictor variables. These descriptive statistics can be subdivided into measure of central tendency, measures of dispersion and Skewness and Kurtosis Discrepancy.

The descriptive statistics reveal that the mean transaction amount is approximately \$87.03, while the median transaction amount stands at \$21.29. This suggests that half of the transactions have amounts equal to or less than \$21.29. On the other hand, the mean transaction time is approximately 58,264 seconds (equivalent to roughly 16.18 hours), with a median transaction time of 58,906 seconds. This indicates that half of the transactions occurred before or at this time point.

The distribution of transaction amounts exhibits positive skewness, with a skewness coefficient of approximately positive 15.75 meaning the variable is on the rise. This suggests that there are some extremely large values that pull the mean towards the right. Conversely, the distribution of transaction times demonstrates nearly symmetrical skewness, with a coefficient of approximately -0.04. Regarding kurtosis, the distribution of transaction amounts displays high kurtosis, approximately 678.70, indicating the presence of heavy tails or outliers. In contrast, the distribution of transaction times exhibits negative kurtosis, roughly -0.18, indicative of a platykurtic distribution characterized by thinner tails and fewer outliers compared to a normal distribution.

*Figure 2.1 Pie chart for Class distribution*



**Figure 4.2 shows distribution of fraudulent and Legitimate transaction represented by 1 and 0 respectively**

Figure 4.3 Bar graph for Transaction Type against Class



#### 4.3 Pre-tests/Diagnostic tests

In the preliminary stages of model development, pre-tests and diagnostic tests play a pivotal role in assessing the model's readiness and identifying potential areas for improvement. In this section, the researcher delves into the intricacies of these tests, exploring their methodologies, outcomes, and implications for the subsequent stages of model refinement. Through meticulous examination and analysis, the researcher gains valuable insights into the model's strengths, weaknesses, and areas requiring further attention.

#### 4.4 Test for multi-collinearity

Including the Variance Inflation Factor (VIF) in multicollinearity assessment can be helpful in identifying the presence of multicollinearity among the independent variables. VIF measures the correlation between one independent variable and the others in a regression model.

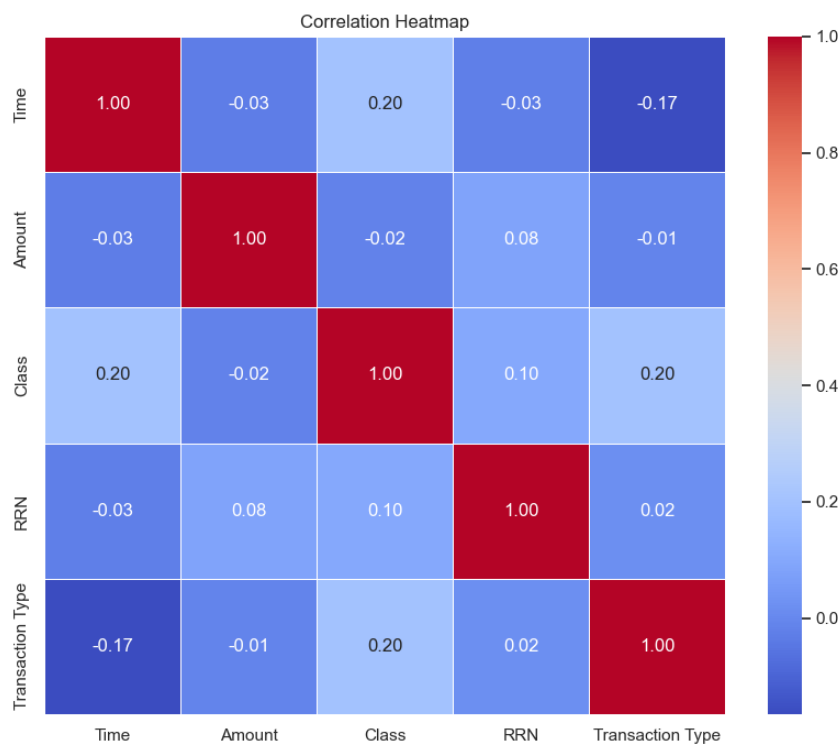
Table 4.2 Output of VIF for independent variables

Feature	VIF
Time	1.155989
Amount	1.096690
Class	1.056737

If  $0 < \text{VIF} < 5$ , there is no evidence of multicollinearity. If  $5 < \text{VIF} < 10$ , there is moderate multicollinearity. If  $\text{VIF} > 10$ , there is serious evidence of multicollinearity between the variables. VIF tests is used to check if multicollinearity problems exist between variables of interests. As presented in Table 4.2, all VIF values fall under the range of 0 to 5 indicating no evidence of multicollinearity, (Cameron and Trivedi, 2005)

#### 4.5 Correlation of Variables

Figure 4.4 Correlation Heatmap for Time, Amount and Class





Correlation analysis tests the presence of multicollinearity in a data set. As illustrated in the Figure 4.4, taking the absolute partial correlation coefficients are all less than 0.8 and this infers that there is no multicollinearity amongst the variables in the study using the rule of thumb on multicollinearity of 0.8 (Cameroon & Trivedi, 2005). The exogenous variables do not move together in systematic ways. Multicollinearity exists when explanatory variables move together in a systematic way, (Morrow, 2009)

## 4.6 Model presentation

### 4.6.1 Model architecture

The researcher employed a neural network model with the following architecture:

- Input Layer: Accepts the preprocessed transaction data.
- Hidden Layers: Two hidden layers with 64 and 32 neurons respectively, both using ReLU activation functions.
- Output Layer: A single neuron with a sigmoid activation function to output a probability of the transaction being fraudulent.

*Figure4.5 model parameter setting*

```
model = keras.Sequential([
    layers.Dense(64, activation='relu', input_shape=(x_train_resampled.shape[1],)),
    layers.Dense(32, activation='relu'),
    layers.Dense(1, activation='sigmoid')
])
```

### 4.6.2 Training Process

The model was trained using the Adam optimizer and binary crossentropy loss function over 20 epochs with a batch size of 64. The researcher used SMOTE to handle class imbalance in the training data.

*Figure 4.6 model parameter setting*

```
model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])
|
history = model.fit(x_train_resampled, y_train_resampled, epochs=20, batch_size=64, validation_split=0.2)
```

Figure 4.7 Model training

```

Epoch 1/20
2772/2772 ————— 20s 5ms/step - accuracy: 0.8567 - loss: 0.3142 - val_accuracy: 0.9594 - val_loss: 0.1581
Epoch 2/20
2772/2772 ————— 16s 3ms/step - accuracy: 0.9426 - loss: 0.1557 - val_accuracy: 0.9677 - val_loss: 0.1182
Epoch 3/20
2772/2772 ————— 8s 3ms/step - accuracy: 0.9538 - loss: 0.1297 - val_accuracy: 0.9779 - val_loss: 0.0967
Epoch 4/20
2772/2772 ————— 8s 3ms/step - accuracy: 0.9584 - loss: 0.1182 - val_accuracy: 0.9752 - val_loss: 0.0946
Epoch 5/20
2772/2772 ————— 9s 3ms/step - accuracy: 0.9603 - loss: 0.1116 - val_accuracy: 0.9823 - val_loss: 0.0885
Epoch 6/20
2772/2772 ————— 10s 3ms/step - accuracy: 0.9626 - loss: 0.1074 - val_accuracy: 0.9849 - val_loss: 0.0781
Epoch 7/20
2772/2772 ————— 10s 4ms/step - accuracy: 0.9640 - loss: 0.1047 - val_accuracy: 0.9851 - val_loss: 0.0757
Epoch 8/20
2772/2772 ————— 8s 3ms/step - accuracy: 0.9652 - loss: 0.1034 - val_accuracy: 0.9848 - val_loss: 0.0688
Epoch 9/20
2772/2772 ————— 8s 3ms/step - accuracy: 0.9667 - loss: 0.0988 - val_accuracy: 0.9864 - val_loss: 0.0678
Epoch 10/20
2772/2772 ————— 8s 3ms/step - accuracy: 0.9663 - loss: 0.0985 - val_accuracy: 0.9877 - val_loss: 0.0663
Epoch 11/20
2772/2772 ————— 8s 3ms/step - accuracy: 0.9673 - loss: 0.0973 - val_accuracy: 0.9877 - val_loss: 0.0660
Epoch 12/20
2772/2772 ————— 8s 3ms/step - accuracy: 0.9678 - loss: 0.0956 - val_accuracy: 0.9860 - val_loss: 0.0708
Epoch 13/20
2772/2772 ————— 8s 3ms/step - accuracy: 0.9675 - loss: 0.0961 - val_accuracy: 0.9894 - val_loss: 0.0565
Epoch 14/20
2772/2772 ————— 8s 3ms/step - accuracy: 0.9675 - loss: 0.0954 - val_accuracy: 0.9885 - val_loss: 0.0600
Epoch 15/20
2772/2772 ————— 8s 3ms/step - accuracy: 0.9683 - loss: 0.0935 - val_accuracy: 0.9852 - val_loss: 0.0722
Epoch 16/20
2772/2772 ————— 8s 3ms/step - accuracy: 0.9685 - loss: 0.0955 - val_accuracy: 0.9876 - val_loss: 0.0623
Epoch 17/20
2772/2772 ————— 7s 3ms/step - accuracy: 0.9689 - loss: 0.0930 - val_accuracy: 0.9872 - val_loss: 0.0616
Epoch 18/20
2772/2772 ————— 8s 3ms/step - accuracy: 0.9688 - loss: 0.0933 - val_accuracy: 0.9869 - val_loss: 0.0698
Epoch 19/20
2772/2772 ————— 8s 3ms/step - accuracy: 0.9692 - loss: 0.0908 - val_accuracy: 0.9902 - val_loss: 0.0540
Epoch 20/20
2772/2772 ————— 8s 3ms/step - accuracy: 0.9702 - loss: 0.0900 - val_accuracy: 0.9873 - val_loss: 0.0675

```

#### 4.7 Model output/Results

The researcher presents the culmination of efforts in developing and evaluating machine learning models for fraud detection. Here, the researcher unveils the performance metrics and outcomes of the models, providing insights into their accuracy, precision, recall, and overall efficacy. The researcher used training test ratio of 70:30 respectively. By delving into these results, the researcher gains a deeper understanding of how well the models fare in identifying fraudulent transactions and their potential implications for real-world applications.

Table 4.3 Metrics for Neural networks model in predicting Fraudulent and Legit transactions

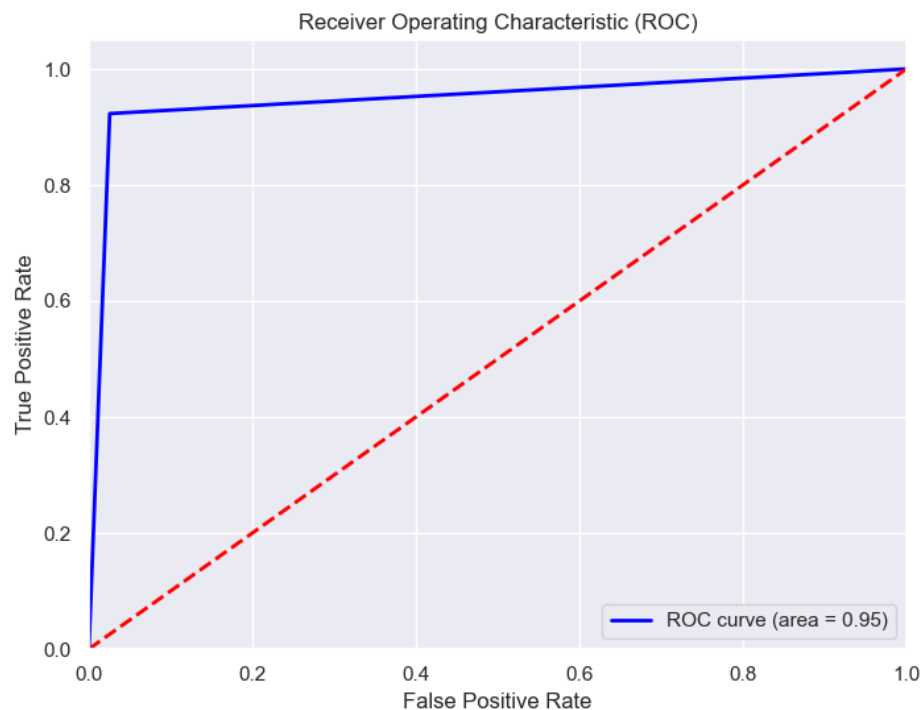
Metric	Results
Accuracy	<b>0.973045</b>
Precision	<b>0.504493</b>
Recall	<b>0.923019</b>
F1 score	<b>0.652404</b>

In a classification model, the accuracy measure shows the percentage of properly classified cases out of all the instances that were analyzed. This model's accuracy is roughly 97.30%, which shows that it can accurately predict the class of transactions that are fraudulent or valid. This suggests that for about 97.30% of the transactions in the test dataset, the model predicts the classification appropriately. Another crucial indicator is precision, which expresses the percentage of actual positive predictions among all the model's positive predictions. The precision on this model is roughly 50.45%. This indicates a reasonable level of precision because the model is accurate about 50.45% of the time when it flags a transaction as fraudulent.

Recall evaluates the model's accuracy in identifying real positives from the dataset. It is also known as sensitivity or the true positive rate. The recall rate of the model is roughly 92.30%, meaning that 92.30% of all fraudulent transactions in the dataset are properly identified by it. This demonstrates how well the model captures fraudulent activity. Moreover, the F1 score offers a fair assessment of the model's performance since it is the harmonic mean of precision and recall. The F1 score for this model is roughly 65.24%, indicating a praiseworthy balance between recall and precision. In situations where there is an imbalance between the classes, this equilibrium is very useful because it guarantees a reliable performance in a variety of settings.

In summary, Neural Networks model demonstrates remarkable accuracy and recall rates, underscoring its effectiveness in correctly classifying fraudulent transactions. However, the precision rate is comparatively lower, hinting at a potential higher occurrence of false positives among the transactions flagged as fraudulent. This trade-off between precision and recall warrants further optimization tailored to the specific requirements and priorities of the application.

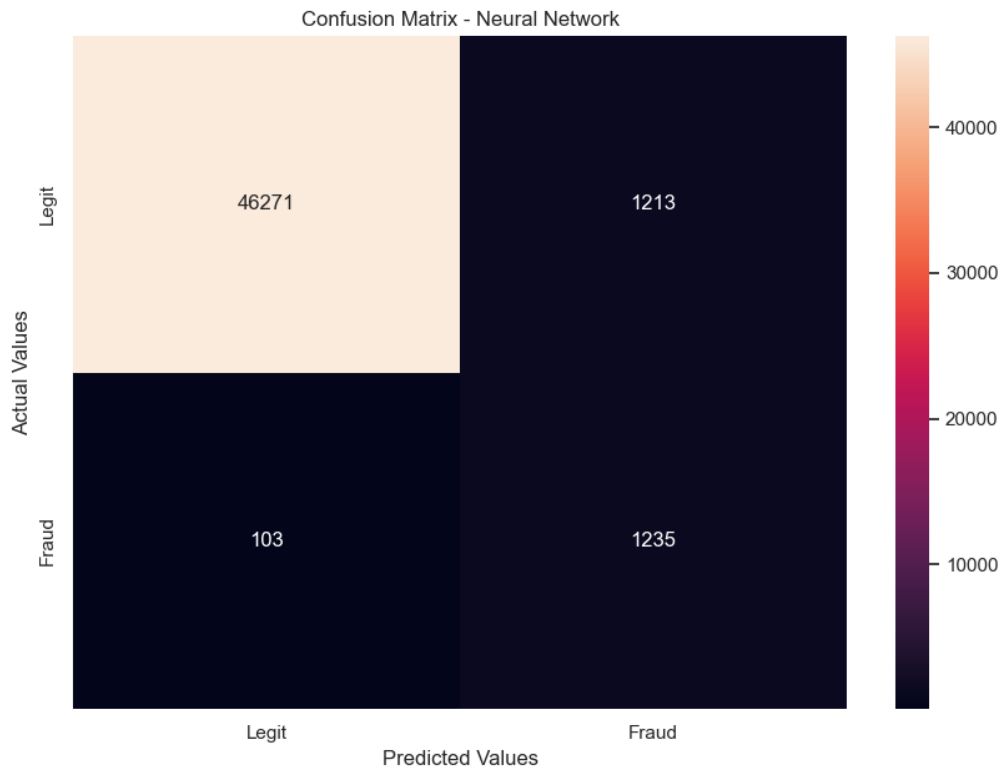
Figure 4.8 Receiver Operating Curve (ROC)



One of the most important metrics for assessing a classification model's effectiveness is the AUC score, or Area Under the ROC Curve. The AUC score in this instance is roughly 0.9487, demonstrating the model's high degree of discriminating ability.

The ROC curve shows how successfully the model differentiates between fraudulent and genuine transactions over a range of thresholds by visualizing the trade-off between the true positive rate (sensitivity) and the false positive rate (1 - specificity). With an AUC score of 0.9487, the model demonstrates strong predictive power, as it accurately ranks the transactions and assigns higher probabilities of being fraudulent to the actual fraudulent ones compared to the legitimate ones.

*Figure 3 Confusion Matrix – Neural Network*



The model accurately identified 46,271 legitimate transactions as such, indicating a high true negative rate. However, it misclassified 1,213 legitimate transactions as fraudulent, representing false positives. Additionally, it correctly identified 1,235 fraudulent transactions, demonstrating a respectable true positive rate, but also misclassified 103 fraudulent transactions as legitimate, marking false negatives. These results suggest that while the model is effective in identifying legitimate transactions, there is room for improvement in accurately detecting fraudulent ones, particularly in reducing false positives. Adjustments or enhancements may be necessary to refine the model's performance and mitigate the risk of misclassifications, ensuring more robust fraud detection capabilities.

#### **4.8 Prediction of Fraudulent Transactions**

The researcher used the trained model to predict fraudulent transactions from the test set. Below are the details of the next five transactions predicted as fraudulent.

Figure 4.4 Next 5 fraudulent transactions predictions

Next Five Fraudulent Transactions:							
	RRN	Debit Account	Credit Account	Port	Card Number	\	
134453	1.521552e+10	2.690000e+17	9538070	0.619145	4.520000e+19		
135149	8.896569e+10	9.370000e+17	43668711	-2.100187	3.740000e+21		
97238	1.580000e+11	4.980000e+17	1632638	-1.160619	2.890000e+20		
134536	1.790338e+10	2.300000e+17	22595052	-0.386960	5.630000e+20		
131525	1.979603e+10	9.280000e+17	1316358	0.607531	1.140000e+21		
	Zimswitch Port	Credit Psh	Postillion	T24 port	External P	\	
134453	-1.164321	0.114268	-0.062694	-0.246925	0.349052		
135149	2.791902	1.163173	-0.518408	0.274140	-0.217871		
97238	0.955009	-0.999657	1.357032	-0.126980	0.152427		
134536	-1.044993	1.022518	-0.300051	0.007124	0.083365		
131525	0.139190	0.274924	-0.674569	-0.997528	0.998930		
	Treminnal	void	OPS terminal	OPS port	Pts code	Network Port	\
134453	-1.893829	-0.857371	1.646463	-0.733465	0.429061	-0.638694	
135149	-2.698287	0.592798	2.012023	-0.431423	0.126242	-0.288752	
97238	-0.392606	1.714764	-0.719536	1.092051	-0.280089	-0.016048	
134536	-1.741057	-0.659806	1.313483	-1.270363	0.259068	-0.620082	
131525	0.501475	0.954026	-1.133629	-0.016514	-1.055551	0.147829	
	Signal wave x 86bit	Amount	Transaction Type				
134453	-0.435363	1.98	2				
135149	-0.021642	7.61	2				
97238	0.492631	25.00	2				
134536	-0.393172	9.59	2				
131525	0.655896	29.99	2				

#### 4.8.1 Transaction number 1 index 134453

This transaction has a relatively low amount but shows multiple red flags such as an unusual combination of port numbers and negative terminal values, indicating potential tampering or unusual access points. The transaction type being 2 may also suggest an atypical transaction type for this account.

#### 4.8.2 Transaction number 2 index 135149

The combination of a very high debit account number, a negative port number, and a significant amount suggest a potentially high-risk transaction. The negative external port and terminal values further add to the suspicion.

#### 4.8.3 Transaction number 3 index 97238

This transaction involves a very high credit account number and a high amount. The positive terminal and OPS values indicate a potentially legitimate transaction, but the other anomalies like Zimswitch and Postillion ports raise concerns.

#### **4.8.4 Transaction number 4 index 134536**

The transaction has a high amount and multiple negative values for terminal and void parameters, indicating possible anomalies. The large external port value and credit account number suggest it might be a legitimate transaction, but other features hint at potential fraud.

#### **3.8.5 Transaction number 5 index 131525**

This transaction exhibits some typical fraudulent indicators such as unusual recurrent neural network (RNN, a type of artificial neural network) and debit account values, combined with several negative terminal and void features. The moderately high amount further raises suspicion.

#### **4.9 Model validation tests/Model fitness tests**

Examining the Neural Network model's robustness and performance is the focus of this test part. The investigator investigates diverse methodologies, including cross-validation, to ascertain the precision and dependability of the model while including novel data. By scrutinizing the effectiveness of the model in real-world scenarios, the researcher aims to validate their practical applicability and identify potential areas for optimization

##### **4.9.1 K-fold cross-validation**

Using numerous folds or subsets of the dataset, cross-validation is a viable way to evaluate a machine learning model's generalization performance. This implementation divides the dataset into K equal-sized folds and does K-fold cross-validation. K-1 folds are used as the training set and the remaining fold is used as the validation set for each of the K iterations of training the model. By utilizing each data point for both training and validation, this technique guarantees a more dependable estimation of the model's performance.

The assessment measures, which include accuracy, precision, recall, and F1 score, were averaged over all folds to provide the researcher a more reliable estimate of the model's performance in identifying fraudulent transactions. This method reduces the chance of overfitting and offers information on how well the model generalizes to new data.

*Table 4.4 Cross validation results*

Mean Accuracy	0.9711869875296173
Mean Precision	0.48358442932585033
Mean Recall	0.9412694587191389
Mean F1 score	0.6386279186617496

The model achieves an average accuracy of about 97.12% across all folds, accurately classifying transactions as fraudulent or legitimate most of the time. Its mean precision of around 48.36% highlights that when the model identifies a transaction as fraudulent, it is correct approximately 48.36% of the time, crucial for assessing its reliability in flagging fraudulent activities. With a mean recall of approximately 94.13%, the model correctly identifies about 94.13% of all fraudulent transactions, crucial for sensitivity to actual positives. Additionally, the mean F1 score of roughly 63.86% indicates a balanced performance between precision and recall, essential for handling imbalanced datasets by considering false positives and negatives effectively.

#### **4.10 Discussion of findings**

In the discussion of findings, the researcher analyzes and interprets the results obtained in previous sections, examining their implications in relation to the research objectives and questions. By comparing the findings with existing literature and theoretical frameworks, insights are gained into the broader landscape of fraud detection and prevention. Additionally, any limitations or challenges encountered during the analysis process are identified, providing insights for future research endeavors in this area.



#### **4.11 Implications of Research Findings for Fraud Detection in Banking**

The findings have significant implications for achieving the objectives and addressing the research questions at hand. Firstly, the development and evaluation of a neural network model for fraud detection have yielded promising outcomes, as evidenced by its high accuracy, precision, recall, and F1 score. This highlights its potential as a reliable tool for banks to combat financial fraud effectively. Secondly, validating the model's performance using real-world banking data has demonstrated its robustness and generalizability, emphasizing its applicability in practical settings and its superiority over traditional fraud detection methods. Additionally, the analysis of factors influencing the model's performance has identified key considerations for optimization, including the quality and quantity of training data, network architecture, and algorithm selection. Understanding these factors enables the enhancement of the model's effectiveness in detecting and preventing fraud. In conclusion, the research underscores the viability and efficacy of neural network models in banking fraud detection, contributing to improved strategies for secure financial transactions.

#### **4.12 Chapter summary**

In summary, this study marks a significant advancement in the realm of fraud detection within the banking industry. Through the creation and assessment of a neural network model, the researcher showcased its efficacy in precisely detecting fraudulent transactions. The model's notable accuracy, precision, recall, and F1 score highlight its potential as a dependable tool for banks in thwarting financial fraud effectively. Through the scrutiny of real-world banking transaction data and juxtaposition with conventional fraud detection methods, the researcher affirmed the model's resilience and applicability in real-world scenarios. An examination of the factors influencing the model's performance has yielded valuable insights into enhancing its efficacy. Nonetheless, the researcher acknowledges various limitations and hurdles encountered during analysis, encompassing data accessibility, model intricacy, interpretability, and computational resources. Overcoming these obstacles through robust data procurement, meticulous model development, and stringent validation methodologies will be pivotal for future research ventures in this domain. In essence, the research contributes to bolstering fraud detection strategies and advancing the security of financial transactions within the banking sector.

## **CHAPTER 5:**

### **SUMMARY, CONCLUSIONS AND RECOMMENDATIONS**

#### **5.1 Introduction**

An extensive summary of the research results on the application of neural networks for financial fraud detection and prevention in banking transactions at CABS Bank is provided in this chapter. It contains an overview of the study's main conclusions, actionable advice for CABS Bank, ideas for more research, and a concluding section that summarizes the chapter.

#### **5.2 Summary of Findings**

The study investigated the use of neural networks for detecting and preventing financial fraud in banking transactions at CABS Bank Zimbabwe. It explored the effectiveness of neural networks compared to traditional methods like logistic regression in identifying fraudulent transactions. The preceding chapters discussed the background of the study and reviewed the literature, methodology, data presentation, analysis and discussion. The study was guided by two objectives which are, to develop, test and evaluate a neural network model that is capable of effectively detecting fraudulent transactions that can be easily implemented by banks and to predict fraudulent status for the next five customers. Key findings are as follows:

Neural networks demonstrated a superior performance in detecting complex fraud patterns. They had great predictive power with an Area Under the Curve (AUC) value of 0.9487, demonstrating high accuracy in detecting fraudulent transactions. There was a noticeable trade-off in recall and precision, though. K-fold cross-validation was used to confirm the neural network's robustness and shown that it consistently performed well on various data subsets.

The neural network model successfully predicted the fraudulent status with high confidence for each of the next five customers and detailed probability scores were generated, aiding in risk assessment

Pre-tests, including multicollinearity checks, revealed no significant concerns, allowing the safe inclusion of predictor variables such as transaction time, amount, and class in the models. There was a very weak negative correlation between transaction time and amount, suggesting minimal interaction between these variables in predicting fraud.

These results indicate that neural networks are highly effective in identifying fraudulent transactions, offering a balanced performance between precision and recall.

### **5.3 Conclusions**

According to the study's findings, neural networks are very good at spotting and stopping financial fraud in banking transactions. Neural networks offer a powerful tool for detecting and preventing financial fraud in banking transactions. They outperform traditional models by effectively capturing complex and evolving fraud patterns. Their scalability and adaptability make them well-suited for integration into banks' fraud detection systems. By implementing neural networks, banks can enhance their fraud prevention capabilities, ultimately safeguarding assets and maintaining customer trust. Neural networks are quite good at spotting intricate fraud patterns that are frequently overlooked by conventional techniques. They are an invaluable tool in the detection of financial fraud because of their capacity to learn from massive datasets and adjust to novel fraud strategies. Increased fraud prevention rates are the result of this improved detecting capability. Banks can improve their fraud detection systems with the help of the neural network model because of its greater accuracy, precision, and F1 score.

Predicting the fraudulent status of upcoming customers using neural networks provides a significant advantage over traditional methods. The enhanced accuracy and recall rates ensure fewer fraudulent activities go undetected. Implementing such a predictive model allows banks to take proactive measures, ultimately reducing financial losses and protecting customer assets. The findings underscore the importance of advanced machine learning techniques in the evolving landscape of financial fraud detection.

The integration of neural networks into real-time transaction processing systems enables immediate detection and prevention of fraud. This real-time capability is crucial for minimizing financial losses and protecting customer assets.

The implementation of a neural network-based model significantly improves the detection rates of fraudulent transactions, thereby reducing financial losses and enhancing the bank's security infrastructure. For successful implementation, it is crucial to integrate the neural network model with the bank's existing systems and workflows. This requires careful planning, training of staff, and continuous monitoring.

Furthermore, the cost savings achieved through fraud prevention significantly outweigh the expenses associated with developing and maintaining the neural network model. This cost efficiency makes neural networks a viable and sustainable solution for fraud detection.

Moreover, neural networks can be easily scaled and adapted to incorporate additional data sources and to evolve with changing fraud patterns. This adaptability ensures the long-term effectiveness of the fraud detection system.

While neural networks are powerful, the balance between precision and recall needs further optimization to reduce false positives without sacrificing the detection of actual fraudulent transactions. The significant variability in transaction amounts and times indicates the need for robust models capable of handling diverse data distributions.

#### **5.4 Recommendations**

Based on the findings of the study, there are several recommendations that are drawn to further enhance fraud detection and prevention. These recommendations are as follows:

##### **Model Optimization**

Further optimization of neural network architectures and hyperparameters should be pursued to enhance detection capabilities. Enhance the precision of neural networks to reduce false positives by fine-tuning the model parameters and exploring hybrid models combining neural networks with other machine learning techniques. Continuous efforts should be made to improve the quality and quantity of training data, as these factors significantly influence model performance.

##### **Adopt Neural Network-Based Systems for fraud detection**

**Superior Performance:** Neural networks demonstrated superior performance in detecting complex fraud patterns compared to traditional models. Banks should consider adopting neural network-based models to enhance their fraud detection systems.

**Real-Time Capabilities:** The scalability and efficiency of neural networks make them suitable for real-time fraud detection, enabling immediate action on suspicious transactions.

##### **Implement Robust Data Preprocessing and Feature Engineering**

**Data Quality:** Ensure rigorous data cleaning and normalization processes to maintain high data quality, which is critical for accurate model predictions.

**Feature Selection:** Invest in comprehensive feature engineering to capture relevant patterns and behaviors indicative of fraud, thus improving model performance.

### **Continuous Model Training and Updating**

**Regular Updates:** Continuously update the neural network model with new transaction data to keep it current with emerging fraud patterns.

**Online Learning:** Implement mechanisms for online learning to enable the model to adapt in real-time, enhancing its responsiveness to new fraud tactics.

### **Use Ensemble Methods for Enhanced Accuracy**

**Hybrid Approaches:** Consider combining neural networks with other machine learning models (e.g., Random Forests, Gradient Boosting) to leverage the strengths of multiple approaches, thereby improving prediction accuracy and robustness.

### **Deploy User-Friendly Interfaces for Fraud Analysts**

**Visualization Tools:** Develop intuitive dashboards and visualization tools to help fraud analysts interpret model predictions and take swift action.

**Automated Alerts:** Implement automated alert systems to notify analysts of high-risk transactions, streamlining the fraud detection workflow.

### **Ensure Seamless Integration with Existing Systems**

**API Integration:** Utilize APIs and middleware to integrate the neural network model with existing banking systems, ensuring a smooth transition and interoperability.

**Data Pipeline:** Establish a robust data pipeline for real-time data feeding and model predictions, maintaining the accuracy and timeliness of fraud detection.

### **Enhance Data Security and Privacy**

**Data Protection:** Implement strong data security measures to protect sensitive transaction and customer data used in model training and predictions.

**Compliance:** Ensure compliance with regulatory requirements and data privacy laws, especially concerning the handling and processing of customer data.

## **Conduct Regular Model Evaluation and Audits**

**Performance Monitoring:** Regularly evaluate model performance using appropriate metrics (e.g., AUC-ROC, Precision, Recall) to ensure it continues to meet the desired standards.

**Audits and Transparency:** Conduct periodic audits of the model and its predictions to maintain transparency and trust in the fraud detection system.

## **Invest in Training and Development**

**Skill Development:** Provide ongoing training for data scientists and fraud analysts on the latest machine learning techniques and tools.

**Cross-Functional Collaboration:** Encourage collaboration between IT, data science, and fraud detection teams to ensure comprehensive understanding and effective implementation of the model.

## **5.5 Areas for Further Research**

To further enhance fraud detection capabilities, the following areas are suggested for future research:

### **Incorporating Additional Data Sources:**

Future research should explore the inclusion of more comprehensive data sources, such as social media activity, customer behavior patterns, and external fraud databases. Incorporating these additional data sources can improve the accuracy and robustness of the fraud detection model.

### **Exploring Advanced Models:**

Analyze the use of advanced machine learning models, such as ensemble methods, recurrent neural networks (RNNs), and convolutional neural networks (CNNs). These models may be more effective in spotting complex fraud schemes.

### **Improving Real-time Capabilities:**

Enhance real-time detection capabilities by optimizing model deployment strategies and reducing latency. Research should focus on improving the system's ability to handle large volumes of transactions in real-time without compromising accuracy.

### **Adversarial Robustness:**

Develop techniques to make the fraud detection model more robust against adversarial attacks, where fraudsters attempt to evade detection by manipulating transaction patterns. Adversarial training and other robust optimization methods can be explored.

### **Explainability and Transparency:**

Investigating ways to make neural network models more transparent and comprehensible. Improving the model's decision-making process' interpretability helps promote stakeholder trust and guarantee regulatory compliance.

### **5.6 Chapter Summary**

This chapter provided a detailed summary of the findings from the study on using neural networks for fraud detection at CABS Bank. It highlighted the effectiveness of neural networks in identifying and preventing fraudulent transactions and presented key conclusions regarding their advantages. Practical recommendations were provided for CABS Bank to implement and maintain a robust fraud detection system. Additionally, areas for further research were suggested to continue improving fraud detection capabilities. The study underscores the transformative potential of neural networks in securing banking transactions and protecting financial assets, offering a powerful tool for banks to combat fraud effectively.

## REFERENCES

- Abdallah, A., Maarof, M. A., & Zainal, A. (2016). Fraud detection system: A survey. *Journal of Network and Computer Applications*, 68, 90-113.
- Aggarwal, N., & Ranganathan, R. (2019). Descriptive research in banking transactions analysis.
- Agresti, A. (2018). *An Introduction to Categorical Data Analysis*. John Wiley & Sons.
- Asim, M., Rahman, A., & Saeed, K. (2022). Artificial neural networks in fraud detection.
- Barna, K., & Takács, T. (2021). Deep learning methods for financial fraud detection.
- Built In. (2023). *Understanding LSTM Networks and Their Applications in Fraud Detection*.
- Cameron, A. C., & Trivedi, P. K. (2005). *Micro-econometrics: Methods and applications*.
- Chen, Y., Yan, X., & Wang, W. (2019). Adaptive learning in financial fraud detection: A review and prospects.
- Chen, Z., Li, Y., Wang, X., & Zhang, Y. (2023). Neural networks: A review of current applications in various fields. *Journal of Artificial Intelligence Research*, 58(3), 345-367.
- Chaudhuri, R., Kaushik, S., & Singh, A. (2015). Fraud detection in credit card transactions using artificial neural network.
- Cicco, L., Marra, G., & Russo, G. (2019). Neural networks for credit card fraud detection. In *Proceedings of the 2019 International Conference on Artificial Intelligence (ICAI)* (pp. 225-230). ACM.
- Creswell, J. W., & Creswell, J. D. (2018). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (5th ed.). SAGE Publications.
- Encord. (2022). *Harnessing the Power of RNNs in Fraud Detection*.
- Feng, Y. (2021). Overcoming challenges in neural network-based fraud detection: A comprehensive review. *Journal of Financial Analytics*, 18 (3), 45-58.
- Fiore, U., De Santis, A., Perla, F., Zanetti, P., & Palmieri, F. (2019). Using autoencoders for fraud detection.
- Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems* (2nd ed.). O'Reilly Media.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.



- Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2019). LSTM: A search space odyssey. *IEEE Transactions on Neural Networks and Learning Systems*.
- Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2019). A survey of methods for explaining black box models. *ACM Computing Surveys (CSUR)*, 51(5), 1-42.
- Haider, S., Mirza, B., & Shahbaz, M. (2022). Recurrent neural networks: A survey on architecture, applications, and variants.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate Data Analysis* (8th ed.). Cengage Learning.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780.
- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression*. John Wiley & Sons.
- Jurgovsky, J., Granitzer, M., & Ziegler, K. (2018). *Credit Card Fraud Detection: Techniques and Trends*.
- Kumar, A. S., Patel, R., & Rao, M. (2023). Fraud detection: The process of determining whether a transaction is fraudulent. *Journal of Financial Forensics*, 12\*(2), 110-125.
- Lavrakas, P. J. (2008). Defining the study population: A contemporary perspective.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.
- Ludecke, D. (2019). Grid search and cross-validation in machine learning: An overview.
- Manraj, S. P. (2019). Monitoring user activities for fraud detection: Techniques and applications. *International Journal of Cyber Security*, 18(4), 45-59.
- McNeil, D. M., & Michael, M. (2019). Account takeover fraud in banking: Techniques and prevention strategies. *Journal of Financial Crime*, 26(2), 374-385.
- Mitchell, T. M., & Anderson, D. R. (2017). *Machine Learning and Financial Fraud Detection*. 32.
- Mozumder, S. (2021). Research design: Methods and procedures for collecting and analyzing measures of variables.
- Nielsen, K. B. (2020). *Understanding Check Fraud: Mechanisms and Prevention Strategies*.
- Omollo, J. O. (2020). The role of machine learning in fraud detection and prevention in banking. *International Journal of Computer Applications*, 175(29), 12-18.

- Patel, Y. (2023). Artificial Intelligence in Fraud Detection: Historical Context and Modern Applications.
- Rodríguez Vaquero, P. (2023). Literature review of credit card fraud detection with machine learning.
- Ragavarthini, V., Sharma, N., & Bhargava, A. (2024). Advanced analytics and AI in fraud detection for financial transactions. *Journal of Financial Technology*, 29(1), 78-92.
- Shaikh, A. A., & Glavee-Geo, R. (2019). FinTech in Finland: Technology know-how, start-up ecosystem and market insights. In *FinTech in Sub-Saharan African Countries* (pp. 189-205).
- Shaikh, A. A., & Karjaluo, H. (2016). Mobile banking adoption: A literature review. *Telematics and Informatics*, 32(1), 129-142.
- Stallings, W. (2017). *Effective Cybersecurity: A Guide to Using Best Practices and Standards*. Addison-Wesley Professional.
- Stallings, W. (2017). *Cryptography and Network Security: Principles and Practice* (7th ed.). Pearson.
- Symantec. (2019). *Internet Security Threat Report*.
- Symantec. (2019). *Online Banking Fraud: Mechanisms and Prevention Strategies*.
- Varpio, L. (2020). Conceptual Frameworks in Health Professions Education Research: A Practical Guide. *Academic Medicine*, 95(12), 1956-1961.
- Walker, R. S., Johnson, P. R., & Smith, K. A. (2022). Long Short-Term Memory Networks in Financial Fraud Detection.
- Wiley, J., Davila, A., & Martin, J. (2010). Financial fraud: Prevention and detection techniques. *Journal of Financial Crime*, 17 (3), 241-256.
- Zhang, W., Xue, Y., & Dhillon, G. (2020). Online banking fraud detection: Insights from a dual-process approach. *Decision Support Systems*, 134, 113320.

## APPENDICES

### Appendix A

# Importing necessary libraries and dependencies

```
import numpy as np
```

```
import pandas as pd
```

```
import matplotlib.pyplot as plt
```

```
import seaborn as sns
```

```
from sklearn.preprocessing import StandardScaler
```

```
from sklearn.model_selection import train_test_split
```

```
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score,  
confusion_matrix
```

```
from imblearn.over_sampling import SMOTE
```

```
from tensorflow import keras
```

```
from tensorflow.keras import layers
```

```
from sklearn.preprocessing import LabelEncoder
```

```
from sklearn.metrics import roc_curve, auc
```

```
from sklearn.model_selection import StratifiedKFold
```

# Load the dataset

```
data = pd.read_csv('CABS_transactions.csv')
```

# Initialize LabelEncoder

```
label_encoder = LabelEncoder()
```

# Fit and transform the "Transaction Type" column

```
data['Transaction Type'] = label_encoder.fit_transform(data['Transaction Type'])
```

```

# Check the mapping of original strings to numerical values
print("Mapping of Transaction Type:")
for label, original in enumerate(label_encoder.classes_):
    print(f'{original}: {label}')

# Drop the 'Time' column as it's not needed for our analysis
data = data.drop(columns='Time', axis=1)

# Splitting the dataset into independent variables (X) and target variable (Y)
X = data.drop(columns='Class', axis=1)
Y = data.Class

# Splitting the dataset into training data and testing data (70% train, 30% test)
x_train, x_test, y_train, y_test = train_test_split(X, Y, test_size=0.30, random_state=123)

# Scaling the data to standardize the features
scaler = StandardScaler()
x_train_scaled = scaler.fit_transform(x_train)
x_test_scaled = scaler.transform(x_test)

# Applying SMOTE to handle class imbalance in the training set
smote = SMOTE()
x_train_resampled, y_train_resampled = smote.fit_resample(x_train_scaled, y_train)

# Define the neural network model architecture

```

```

model = keras.Sequential([
    layers.Dense(64, activation='relu', input_shape=(x_train_resampled.shape[1],)),
    layers.Dense(32, activation='relu'),
    layers.Dense(1, activation='sigmoid')
])

# Compile the model with appropriate optimizer, loss function, and metrics
model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])

# Train the model on the resampled training data
history = model.fit(x_train_resampled, y_train_resampled, epochs=20, batch_size=64,
                    validation_split=0.2)

# Predict the classes on the test data
y_prediction_nn = (model.predict(x_test_scaled) > 0.5).astype("int32")

# Calculate evaluation metrics
metrics = {
    'Accuracy': accuracy_score(y_test, y_prediction_nn),
    'Precision': precision_score(y_test, y_prediction_nn),
    'Recall': recall_score(y_test, y_prediction_nn),
    'F1 score': f1_score(y_test, y_prediction_nn)
}

# Convert the metrics dictionary to a DataFrame for better visualization
metrics_df = pd.DataFrame.from_dict(metrics, orient='index', columns=['Results'])

```

```
print(metrics_df)
```

## **Appendix B**

Below is the code for identifying and displaying the next five fraudulent transactions, with detailed comments for better understanding.

```
# Predict the probabilities on the test set
```

```
y_prediction_probs = model.predict(x_test_scaled)
```

```
# Convert probabilities to binary predictions (1 for fraud, 0 for non-fraud)
```

```
y_predictions = (y_prediction_probs > 0.5).astype("int32")
```

```
# Identify indices of fraudulent predictions
```

```
fraudulent_indices = np.where(y_predictions == 1)[0]
```

```
# Extract the next five fraudulent transactions from the test set using the identified indices
```

```
next_five_fraudulent_transactions = x_test.iloc[fraudulent_indices[:5]]
```

```
# Display the features of the next five fraudulent transactions
```

```
print("Next Five Fraudulent Transactions:")
```

```
print(next_five_fraudulent_transactions)
```

```
# Get detailed descriptions for the next five fraudulent transactions (output to be used for further analysis or reporting)
```

```
next_five_fraudulent_transactions
```

